Biological Testing

Gail Goodman Snitkoff, PhD

Biological testing includes the quantitative assay of drugs by biological methods as well as the application of qualitative biological tests. Such testing uses intact animals, animal preparations, isolated living tissues and cells, or microorganisms. In addition to drugs, biological tests are a requirement for plastics to be used as containers or closures for ophthalmic and parenteral preparations or to be used as implants, devices, or other related systems.

The practices of the USP are a good index of the state of biological testing. Currently there is a trend to use fewer animals in research and biological testing and to use alternatives such as cells and microorganisms in culture. This decrease in animal use can be observed in the decreased requirement for animal testing by the USP as documented in their monographs. Wherever possible, *in vitro* procedures should be used to complement or replace *in vivo* tests for evaluating the suitability of plastics.

The majority of currently available therapeutic agents are substances of known chemical composition that can be assayed by quantitative chemical or physical analyses. However, there are a limited number of useful drugs that cannot be assayed satisfactorily by chemical or physical means. Such drugs, which are primarily of natural origin, are assayed by biological methods. Biological standardization procedures generally are less precise, more time-consuming, and more expensive to conduct than are chemical assays; therefore, they generally are reserved for use:

1. If the chemical identity of the active principle has not been elucidated fully.
2. If no adequate chemical assay has been devised for the active principle, although its chemical structure has been established (eg, insulin).
3. If the drug is composed of a complex mixture of substances of varying structure and activity (eg, digitalis, posterior pituitary).
4. If purification of the crude drug, sufficient for the performance of a chemical assay, is not possible or practical (eg, the separation of vitamin D from certain irradiated oils).
5. If the chemical assay is not a valid indication of biological activity, due, for example, to lack of differentiation between active and inactive isomers.

There are several situations in which factors such as specificity, sensitivity, or practicality dictate the use of a biological rather than a chemical assay procedure.

A chemical assay quantitatively determines the amount of a specific compound or structural moiety present in a given sample. Once the concentration has been established an assumption is made relative to the biological activity of the sample. In contrast, a biological assay measures the actual biological activity of a given sample, which may represent the algebraic sum of the interaction of a number of chemical and physical-chemical factors. For example, the data obtained from a chemical assay will not provide information concerning the contribution to the net biological activity of trace amounts of substances that do not influence the chemical analysis. Such substances may produce qualitative variations in biological activity that may be responsible for unexpected side effects or toxic reactions. Furthermore, the enhancing or inhibiting influences of variations in the physical state of the active principle are not reflected in the results of a chemical assay. The safety, efficacy, and dependability of dosage of drugs are contingent upon standardization, and biological assays must be employed in some instances even though the chemical identities of the active principles in the preparation may be known.

ANIMAL TESTING

As animals are an important *unknown* factor in most biological assays, the need for their proper selection and adequate care is self-evident. Most laboratories seek a reliable source of animals that can supply their needs from colonies maintained for this purpose. In any one test it is desirable to use animals of only one strain. Usually bioassayists adopt a specified strain for all work of a particular type. This enables the bioassayist to gain experience concerning the expected normal variation. For some assays a specific sex must be employed (eg, estrogenic tests); in other assays either sex may be used, but the effect that sex may play in the response should not be overlooked. The male rat, for instance, has a faster growth rate than the female; therefore, indiscriminate use of both males and females in a rat growth test should be avoided. Differences in the response of the sexes may extend into other categories, such as response toward toxic materials. Animals used in these biological assays should always be handled according to the National Institutes of Health guidelines.[1]

BIOASSAY PROCEDURES

Bioassays are conducted by determining the amount of a preparation of unknown potency required to produce a definite effect on suitable test animals or organs under standard conditions.

REFERENCE STANDARDS

To minimize the source of error resulting from animal variation, standard reference preparations are used in certain bioassay procedures. The principle of the using a reference standard consists of successively testing the unknown and standard preparations on two groups of similar animals, or, in some cases (eg, epinephrine, posterior pituitary) on the same animal or organ. The amount of the unknown preparation required to

produce an effect equal to that produced by a defined amount of the standard will be inversely proportional to their relative potencies. The potency of the unknown therefore can be expressed as a percentage of that of the standard.

In some assays it is necessary to adopt precise methods of calculating potency based upon observations of relative, but not necessarily equal, effects. Likewise, methods of computation have been devised to determine the statistical reliability of the results. These procedures are discussed in Chapter 12. The section on *Biological Tests and Assays* in the USP also presents a detailed consideration of factors germane to the *Design and Analysis of Biological Assays.*

When reference standards are required for use in assays, they are available as a service from USP-NF Reference Standards, 12601 Twinbrook Parkway, Rockville, MD 20852. These references are standardized in terms of the appropriate existing international standards.

DISADVANTAGES OF BIOASSAYS

Biological assays leave much to be desired in several respects. Although some are extremely sensitive in detecting small differences in concentration, their quantitative accuracy usually falls considerably below that obtainable with most chemical analyses. The techniques and interpretations involved often can vary with different operators, in spite of the rigid require-

ments specified by the official publications; hence, there is a considerable subjective element present.

Furthermore, the effect measured in the test animals often is not that which the drug is intended to produce in treating patients. The importance of this discrepancy was minimized formerly, but recent studies have shown that when several active principles are present in a crude drug, those producing the maximal therapeutic effect are not necessarily the ones chiefly responsible for the action measured in the assay. As a result, samples found to be of equal strength by assay may show different potencies when employed clinically. An example of this situation is found in the discussion on digitalis.

CLASSIFICATION OF BIOASSAY PROCEDURES

Bioassays are classified in three groups according to whether the effect produced is all or none (as death), graded (as rise in blood pressure), or is characterized by developing in a measured period of time (as the curative response to thiamine). It should be noted that in all three types, with few exceptions, the calculations of potency are based on the sizes of doses necessary to produce approximately equal effects and not on the intensities of the responses. Furthermore, *the results derived from all are quantitative* in that the potency of the unknown is expressed in terms of the standard.

ANIMAL ASSAYS

In the following section biological assay procedures involving the use of intact animals, animal preparations, or isolated, surviving animal tissues or organs are considered. The presentation is restricted largely to the general principles and basic experimental approaches involved in each of several representative types of biological assay methods. For complete details of the official procedures, the reader should consult the corresponding official compendial monographs.

DIGITALOID DRUGS

The digitaloid group of drugs includes digitalis (the dried leaf of *Digitalis purpurea*), which is used medicinally as the powdered material in the form of capsules or tablets. These products of natural origin contain, in addition to the cardioactive glycosides *digitoxin* and *gitoxin,* a saponin-like glycoside, termed *digitonin,* that is largely devoid of the cardiac effects of digitalis, and a complex mixture of constituents including digitoflavin, digitophyllin, lipids, carbohydrates, and other nonspecific plant components. Although the cardioactive glycosides are similar in chemical structure and pharmacodynamic activity, they differ markedly in milligram potency, efficiency of gastrointestinal (GI) absorption, speed of onset, and duration of action. Furthermore, there is considerable variation among different lots of crude drug in respect to the total active glycoside content and the relative concentration of each active principle.

It is apparent that *chemical* assay procedures, such as the determination of total glycosides or total aglycones, cannot adequately measure the pharmacodynamic activity of the crude drug or galenical preparation of digitalis. Such drugs, composed of a complex mixture of substances of varying structure and activity, must be subjected to a *biological* assay.

The biological assay of digitalis is based on determination of the amount of test material required to cause death due to cardiac arrest in the anesthetized pigeon, relative to the amount of

a reference standard preparation required to produce the same effect.

As a rule, it is desirable that the parameters of a biological assay simulate, as closely as practicable, the conditions generally associated with clinical usage of the drug in question. In this respect the currently employed biological assay procedure for digitalis has several disadvantages and limitations. For example, the preparation to be assayed is given by fractional, intermittent, intravenous injection, whereas in the treatment of patients with cardiac disorders digitaloid drugs most often are administered orally. The assay procedure does not, therefore, take into consideration variations in clinical effectiveness among different members of this medicinal group that may be attributable to differences in the rate and completeness of absorption from the GI tract. Furthermore, the calculation of potency is based on the amount of drug required to produce death of the test animal due to cardiac arrest. Thus the end point of the assay corresponds to a toxic rather than a therapeutically desirable event.

However, it may be claimed that the toxic effects of digitalis on the heart constitute extensions of the cardiodynamic changes that are beneficial in certain cardiac disorders. Because of the limitations of the pigeon method, many attempts have been made to develop more satisfactory bioassay procedures. None of the alternatives devised to date has any compelling advantage over the current official method.

Separation and identification of the active principles of crude drugs having characteristic digitalis-like effects on the heart have resulted in the availability, for therapeutic use, of products consisting essentially of a single relatively pure cardiotonic glycoside. These additional members of the digitaloid group of drugs include Acetyldigitoxin, Deslanoside, Digoxin, and Lanatoside C (all of which are derived from *Digitalis lanata*), Digitoxin (a glycosidal constituent of both *Digitalis purpurea* and *Digitalis lanata*), and Ouabain (a glycoside obtained from the seeds of *Strophanthus gratus*). Preparations

containing these glycosides are assayed quantitatively by spectrometric methods described in Chapter 28.

Chemical assay procedures enable precise determination of the amount of glycoside present in a particular dosage formulation. However, it must be emphasized that the response to digitaloid drugs varies considerably among cardiac patients. A clinical assay must, therefore, be performed with each patient, regardless of whether the digitaloid preparation being used was standardized on the basis of a chemical or biological assay procedure.

Biological assay by the pigeon method is specified for the following official preparations: *Digitalis, Powdered Digitalis, Digitalis Capsules,* and *Digitalis Tablets.*

INSULIN

Insulin is a hormone that is synthesized and secreted by the beta cells of the pancreatic islets and is stored in intracellular granules. Human insulin has a molecular weight of 5807 and is composed of two polypeptide chains joined by two disulfide linkages. The shorter A chain is composed of 21 amino acids and the longer B chain is composed of 30 amino acids. In the beta cells, insulin is synthesized from a long single-chain precursor designated as proinsulin. Through proteolysis, 35 amino acids (connecting chain) are cleaved from proinsulin, resulting in the double-chain insulin molecule.

Although the structure of insulin varies from species to species, these variations are slight. The molecular weight for bovine insulin is 5733, and for porcine insulin is 5777. Most of the insulin used for medicinal purposes is human insulin prepared by recombinant DNA technology.

Insulin is used to control blood-glucose levels in people with diabetes mellitus. Diabetes mellitus, which results from an inadequate secretion of insulin from the pancreas, is characterized by hyperglycemia, ketoacidosis, hyperlipemia, azoturia, and ketonemia.

Accurate standardization of insulin preparations is essential because discrepancies of about 10% from the required dose may result in severe adverse reactions in the diabetic patient. Inadequate insulin replacement therapy may be associated with the appearance of any of the characteristic symptoms of diabetes mellitus, including ketoacidosis and diabetic coma; overdosage may result in marked hypoglycemic reactions.

Preparations of insulin are divided into three classes: fast-, intermediate-, and long-acting types. Insulin Injection and Prompt Insulin Zinc Suspension are classified as fast-acting, with an approximate time of onset of 1 hr. The intermediate-acting preparations include Isophane Insulin Suspension, which has a time of onset of approximately 2 hr. Protamine Zinc Insulin Suspension and Extended Insulin Zinc Suspension are classified as the long-acting preparations, with the time of onset being approximately 4 hr.

There are eight different official insulin monographs with different approaches to standardization. In the USP25/NF20, the monographs for Insulin and Insulin Injection no longer specify the biological (rabbit) Insulin Assay, but only an HPLC assay. However, the monographs use the biological assay results for the rubric definition of potency. The biological assay is of importance because it accurately reflects the activity of the preparation in the diabetic patient. Briefly, the assay is based on a comparison between the potencies of the unknown preparation and the standard preparation in lowering the blood sugar level of intact rabbits following subcutaneous injection. The assay is required for the validation of new insulin preparations and the determination of the activity of insulin analogs. It should be noted that the monographs for Insulin Human and Insulin Human Injection have only an HPLC assay and define potency in terms of chromatographically derived Insulin Human Units; however, both monographs use a *biological potency* specification, which requires a biological assay whose results are not reflected in the rubric statement. In addition, all the

monographs define an insulin unit as equivalent to a specific weight of pure insulin.

The remaining four monographs—Isophane Insulin Suspension, Insulin Zinc Suspension, Extended Insulin Zinc Suspension, and Prompt Insulin Zinc Suspension—do not specify an assay but rather define potency in their rubric statements based on their proper preparation from insulin. All eight monographs involve a requirement for testing for *bacterial endotoxins*. These features of the insulin articles are summarized in Table 31-1.

Insulin preparations are subject to the regulations of the Federal Food, Drug, and Cosmetic (FD&C) Act, which requires certification by the Food and Drug Administration (FDA) of each lot marketed. Many of the tests and criteria employed by the FDA follow closely those specified by the official compendia. In addition to the biological assays, certain chemical and bacteriological tests must be performed to meet the requirements for certification by the FDA. The Act (particularly Section 506) and regulations thereunder should be consulted for specific details concerning the steps that must be followed in obtaining such certification.

Insulin Injection is available as solutions containing 40, 100, and 500 USP Insulin Units per mL. Extended Insulin Zinc Suspension, Prompt Insulin Zinc Suspension, Isophane, and Insulin Zinc Suspensions provide 40, 80, or 100 units/mL. A variation of not more than 5% from the labeled potency is permitted.

GLUCAGON

Glucagon is a straight-chain polypeptide hormone comprised of 29 amino acids and has a molecular weight of 3482.78. The polypeptide structure of human glucagon is identical to the structure of bovine and porcine glucagon.

Glucagon, which is synthesized and secreted by the alpha cells of the islets of Langerhans in the pancreas, acts to increase blood-glucose concentrations by stimulating glycogenolysis in the liver. Glucagon increases the synthesis of cyclic adenosine monophosphate (cAMP) in the liver, which in turn activates phosphorylase, the rate-limiting enzyme in the conversion of liver glycogen to glucose. Glucagon for Injection contains glucagon hydrochloride and one or more suitable dry diluents.

Glucagon is used in the treatment of acute hypoglycemic reactions, especially in diabetics with insulin-induced hypoglycemia. The subcutaneous, intramuscular, or intravenous administration of 1 mg of glucagon elicits a rapid increase in blood-glucose concentration. Glucagon also has a spasmolytic effect on the GI tract and, as such, is used in radiographic examinations to relax the intestinal tract.

In the bioassay of Glucagon for Injection, the glucagon test sample and Glucagon Reference Standard, at a specified concentration, are injected into anesthetized cats and the subsequent elevations in blood glucose are monitored over specified time intervals. In this procedure 16-hour fasted cats are anesthetized with a long-acting barbiturate and both femoral veins are exposed surgically. The glucagon samples are injected into one femoral vein and blood samples are taken from the vein on the opposite side for glucose determinations.

One hour after each cat is injected with a specified amount of the Glucagon Reference Standard, an initial blood sample is withdrawn and each cat is randomized for further assessment in a two 4×4 Latin-squares dose regimen. The specific procedures employed in this bioassay and the method for potency calculation are detailed in the USP-NF.

In the USP XXV/NF 20, the current biological assay for glucagon has been modified to use freshly harvested rat hepatocytes. In this assay, the hepatocytes are suspended in an appropriate medium and individual test tubes are treated with either USP Reference Standard glucagon or unknown glucagon preparations. The amount of glucose released from the treated hepatocytes is measured spectrophometrically and the amount

Table 31-1. Summary of Biological Assay Procedures

COMPENDIAL ARTICLE	ACTIVITY ASSAYED	ANIMAL EMPLOYED	ROUTE OF ADMINISTRATION OF TEST MATERIAL	END POINT OF ASSAY	UNITAGE	ADDITIONAL BIOLOGICAL TESTS REQUIRED
Digitalis	Cardiac (cardiotonic) action	Pigeon	IV infusion	Cardiac arrest (death)	100 mg is equivalent to not less than 1 USP Digitalis Unit	—
Insulin injection	Hypoglycemic	Rabbit	SC injection	Reduction of blood glucose level	1 mL is equivalent to 40, 100, or 500 USP Insulin Units (potency not less than 95% and not more than 105% of that stated on the label)	Bacterial endotoxins
Insulin	Hypoglycemic	Rabbit	SC injection	Reduction of blood glucose level	Each mg has a biological potency of not less than 26.5 USP Insulin Units	Bacterial endotoxins
Insulin Human Insulin Human Injection	—	—	—	—	Biological potency (uses Insulin Assay, but rubric definition uses HPLC)	Bacterial endotoxins
Isophane Insulin Suspension	—	—	—	—	Biological Activity of the supernatant liquid (uses Insulin Assay, with modifications)	Bacterial endotoxins
Glucagon for injection	Hyperglycemic	Cat (injected IP with dextrose 16 hr prior to assay)	IV injection	Elevation of blood glucose	—	—
Oxytocin injection Oxytocin Nasal Solution	— Contraction of isolated rat uterus	Rat hepatocytes Rat	In vitro Water bath with addition of oxytocin	Magnitude of uterine contractions	1mg is equivalent to not less than 400 Oxytocin Units	Release of glucose Pressor activity—oxytocin injection must not contain excessive vasopressor activity as determined by elevation of arterial blood pressure following IV injection of the test sample in phenoxybenzamine pretreated rats
Vasopressin injection	Vasopressor	Rat (pretreated with phenoxybenzamine)	Intermittent IV injection	Elevation of arterial blood pressure	1mg is equivalent to not less than 300 Vasopressin Units	Oxytocic activity—vasopressin injection must not contain excessive oxytocic activity as determined by contraction of uterine smooth muscle isolated from the guinea pig
Posterior pituitary injection	Vasodepressor and vasopressor (Each mg of USP Posterior Pituitary Reference Standard represents 2.4 USP units of oxytocic activity and 2.1 USP units of vasopressor activity.)	Refer to assays for Oxytocin Injection and Vasopressin Injection	—	—	1 mL possesses USP Posterior Pituitary activity equivalent to not less than 85% and not more than 120% of the oxytocic activity and vasopressin activity stated on the label in USP Posterior Pituitary Units	—
Corticotropin injection Repository corticotropin injection Sterile corticotropin zinc hydroxide suspension						

Substance	Activity	Test animal	Method	Response measured	Units	Additional requirements
Corticotropin for Injection	Adrenal cortical stimulation	Hypophysectomized rat	SC injection	Reduction of ascorbic acid content of adrenal glands	—	Vasopressin activity—corticotropin injection must not contain excessive vasopressor activity as determined by elevation of arterial blood pressure following IV injection of the test sample in phenoxyben-amine-pretreated rats (this test is not required with Sterile Corticotropin Zinc Hydroxide Suspension)
Bacterial endotoxins Chorionic gonadotropin Chronic gonadotropin for injection	Gonad-stimulating	Female rat	SC injection daily for 3 days	Increase in weight of uterus	1 mg is equivalent to not less than 1500 USP Chorionic Gonadotropin Units	Estrogenic activity—chorionic gonadotropin must not contain excessive estrogenic activity as determined by cytological examination of vaginal smears taken from ovariectomized rats injected SC with the test sample
Bacterial endotoxins Acute toxicity (determined by minimal toxicity in mice injected IV with 1000 USP Chorionic Gonadotropin Units) Heparin sodium Heparin sodium injection Anticoagulant Heparin Solution Heparin Calcium Heparin Calcium Injection	Anticoagulant	Sheep	*In vitro* addition of heparin sodium to blood plasma	Inhibition of clot formation	1 mg is equivalent to not less than 140 USP Heparin Units when derived from intestinal mucosa or other tissues from domesticated foo animals.	Bacterial endotoxins
Heparin Lock Flush Solution Dihydroergotamine Mesylate, Heparin Sodium and Lidocaine Hydrochloride Injection Protamine sulfate Protamine sulfate injection	—	—	—	—	—	
Protamine sulfate for injection	Heparin neutralization	Sheep	*In vitro* addition of protamine sulfate to blood plasma containing known amounts of heparin sodium	Reduction of clotting time of heparinized plasma	1 mg neutralizes not less than 100 USP Units of heparin activity from lung tissue or not less than 100 USP Units of heparin derived from intestinal mucosa	Bacterial endotoxins
Cod liver oil	Antirachitic (vitamin D)	Rachitic rat	Oral feeding (one-half of total dose on day 1; one-half on day 3 or day 4)	Calcification of rachitic metaphysis of radius and tibia	1 g contains not less than 2.125 µg (85 USP Units) of vitamin D	1 g also contains not less than 255 µg (850 USP Units) of vitamin A (assayed by a spectrophotometric method)

of glucagon is determined by interpolation against a standard curve.

PARATHYROID

Parathyroid hormone is responsible for maintaining extracellular calcium ions at a constant concentration in the body. Porcine, bovine, and human parathyroid hormones are linear polypeptide chains of 84 amino acids with molecular weights of approximately 9500. Amino acids 1 to 27 of the *N*-terminal portion of the peptide are associated with biological activity.

The bioassay for Parathyroid Injection, which is found in USP XXI but is not included in the current USP (USP XXV), is based on measuring the increase in serum calcium in dogs. In this assay, the serum calcium levels are determined just prior to, and 16 to 18 hours after, the subcutaneous injection of the dose of Parathyroid Injection. Each mL of Parathyroid Injection possesses a potency of not less than 100 USP Parathyroid Units. One USP Parathyroid Unit represents 1/100th of the amount of Parathyroid Injection required to raise the calcium content of 100 mL of the blood serum of normal dogs, 1 mg within 16 to 18 hours after administration.

Parathyroid Injection is no longer available for clinical use. Parathyroid hormone was used extensively to raise plasma calcium levels in hypocalcemic patients. Presently, this is achieved more safely by the administration of calcium and/or vitamin D.

POSTERIOR PITUITARY, OXYTOCIN, AND VASOPRESSIN

Extracts of the posterior lobe of the neurohypophysis, when injected into responsive animals, may exert a variety of pharmacodynamic effects, including a rise in blood pressure, contraction of uterine smooth muscle (oxytocic effect), an increased renal tubular reabsorption of water (antidiuresis), and milk-ejection (galactokinesis) in the lactating mammary gland. Although there is no conclusive agreement on the number of different hormones elaborated by the neurohypophysis, two distinct active principles have been separated from extracts of this structure. These are *oxytocin,* which possesses primarily oxytocic and galactokinetic activities, and *vasopressin,* which exhibits predominantly pressor and antidiuretic activities. Both of these hormones are nonapeptides; the amino acid sequences of these fractions obtained from several animal species have been determined, and corresponding nonapeptide amides have been synthesized.

Posterior Pituitary Injection, which is prepared from the posterior lobe of the pituitary gland of domestic animals used for food by man, contains both oxytocic and vasopressor principles in varying amounts. Since oxytocin and vasopressin are available in purified form, Posterior Pituitary Injection, which represents a mixture of the active principles, is used relatively infrequently and is no longer included in the current edition of the USP (XXV).

The potency of Oxytocin Injection used to be determined by monitoring the decreases in blood pressure in an anesthetized chicken following intravenous administration of the Oxytocin Injection. Currently oxytocin activity is monitored by measuring the contraction of an isolated rat uterus. The contractile activity of the sample is compared to that of a reference standard. Oxytocin Nasal Solution, assayed by the same biological method as the Injection, is available for use as a lactational stimulant.

Currently, all commercially available preparations of oxytocin are prepared synthetically. Oxytocin Injection for intravenous or intramuscular administration contains not less than 400 USP oxytocin units per milligram. Oxytocin Injection is used to induce labor, control postpartum uterine bleeding, and treat incomplete abortion.

Vasopressin, also known as antidiuretic hormone, exerts antidiuretic activity and is a potent vasopressor. Vasopressin Injection is prepared synthetically or by extraction of the posterior lobe of the pituitary glands of domestic animals. The potency of Vasopressin Injection is determined by monitoring the elevations in blood pressure in a male rat following intravenous administration. Blood pressure elevations are monitored and compared to those obtained with a standard Vasopressin Injection, to determine potency of the assay sample relative to a USP Vasopressin Reference Standard. On a unit basis, the antidiuretic activity of Vasopressin is greater than or equal to 300 USP Vasopressin units per milligram.

Vasopressin Injection contains not less than 90% and not more than 110% of the USP Vasopressin Units stated on the label. Vasopressin Tannate is available commercially and used to prevent or control the symptoms and complications of diabetes insipidus caused by deficiency of endogenous antidiuretic hormone. All commercially available vasopression is prepared synthetically.

CORTICOTROPIN

Corticotropin (or ACTH, adrenocorticotropic hormone) is a polypeptide hormone that is synthesized and secreted by basophilic cells of the adenohypophysis. This hormone is a straight-chain polypeptide comprised of 39 amino acids. Corticotropin stimulates the release of cortisol, corticosterone, and aldosterone from the adrenal cortex. Stimulation of the median eminence of the hypothalamus causes the release of a polypeptide called corticotropin-releasing factor (CRF) into the circulatory system. CRF stimulates the release of corticotropin from the adenohypophysis. Endogenous corticosteroids influence the secretion of corticotropin through a negative feedback loop. Increased circulating levels of corticosteroids exert a negative influence on the adenohypophysis and decrease the secretion of corticotropin.

Corticotropin Injection is a sterile solution of the polypeptide hormone obtained from the pituitary glands of mammals used for food. It possesses the ability to stimulate the release of corticosteroids from the adrenal cortex. Corticotropin for Injection is the sterile, dry material of the polypeptide hormone and is made into a solution with suitable diluents, buffer, and an antimicrobial agent. Corticotropin Injection and Corticotropin for Injection can be administered by the subcutaneous, intramuscular or intravenous routes. Repository Corticotropin Injection is corticotropin in a solution of partially hydrolyzed gelatin and is intended for subcutaneous and intramuscular administration. Sterile Corticotropin Zinc Hydroxide Suspension is a suspension of corticotropin adsorbed on zinc hydroxide and is intended for intramuscular administration.

The Third International Standard for Corticotropin[2] has been adopted as the reference standard for corticotropin. In the biological assay[3] for corticotropin, rats are injected subcutaneously with specified diluted standard solutions and test solutions of corticotropin, 16 to 48 hours after the removal of the hypophysis. Three hours after the injections, the rats are anesthetized, and both adrenal glands of each rat are removed, and cleaned from adhering tissue, weighed and assayed for ascorbic acid content. The methodology involved in the preparation of the standard and test solutions, as well as the exact procedure of the bioassay, ascorbic acid determination and calculations are detailed in the USP. The bioassay for Corticotropin Injection and Corticotropin for Injection, labeled for intravenous administration, is identical to the procedure outlined above with the exception that the preparations are injected intravenously in the rats. For the bioassay of Sterile Corticotropin Zinc Hydroxide Suspension sufficient 0.1 *N* hydrochloric acid is added to the preparation for solubilization prior to being assayed in rats by the subcutaneous method.

Corticotropin Injection and Repository Corticotropin Injection also are assayed for vasopressin activity according to the procedure for Corticotropin Injection and the assay for Vasopressin Injection in the USP. Anesthetized rats are injected with a specified dose of USP Posterior Pituitary Reference Standard at 12- to 15-minute intervals and the blood pressure elevations are monitored and recorded. At the midpoint of the timed injections of the Reference Standard, specified dilutions of the Corticotropin Injection are injected into the rat and the blood pressure response is recorded. The blood pressure elevation observed with the Corticotropin Injection should not exceed the average elevation observed with the Reference Standard before and after the Corticotropin Injection.

CHORIONIC GONADOTROPIN

Chorionic gonadotropin is a gonad-stimulating principle, of placental origin, prepared from the urine of pregnant women. The biological activity of chorionic gonadotropin is essentially identical to that of the luteinizing hormone (interstitial cell-stimulating hormone) of the anterior pituitary. Chorionic gonadotropin is used in sequence with menotropins (human menopausal gonadotropins) in the treatment of infertility in women in whom anovulation is due to low or absent endogenous gonadotropins. Follicular growth and maturation are promoted by initial treatment with menotropins, followed by administration of chorionic gonadotropin to induce ovulation by simulating the normal preovulatory surge of luteinizing hormone.

Chorionic gonadotropin also is used in the treatment of cryptorchism in cases in which there is no apparent anatomical obstruction to descent of the testis. Combined therapy with this hormone and menotropins may promote spermatogenesis in patients with hypogonadotropic eunuchoidism. Diagnostically, chorionic gonadotropin is used to evaluate Leydig cell responsiveness.

Chorionic Gonadotropin for Injection is a sterile, dry mixture of chorionic gonadotropin with suitable diluents and buffers. Biological assay of the preparation is based on the increase in weight of the uterus excised from young female rats sacrificed 2 days after the last of three daily subcutaneous injections of dilutions of the test sample. The response is compared to that obtained in a series of animals similarly treated with USP Chorionic Gonadotropin Reference Standard. The uterotropic effects depend on elaboration of ovarian hormones in response to the gonad-stimulating activity of chorionic gonadotropin. Chorionic Gonadotropin for Injection is satisfactory if it contains not less than 80% and not more than 125% of the potency stated on the label.

It is also necessary to ascertain, biologically, that Chorionic Gonadotropin for Injection meets the requirements of the estrogenic activity test. This is accomplished by examination of vaginal smears taken from ovariectomized rats on each of three successive days following subcutaneous injection of 0.25 mL of chorionic gonadotropin test solution twice a day (morning and afternoon) for 2 days. The requirements of the test are met if the cellular elements in the smears consist of leukocytes and a few nucleated epithelial cells, but no cornified epithelial cells.

HEPARIN

Heparin consists of straight-chain mucopolysaccharides, called glycosaminoglycans, and has an average molecular weight of 15,000. The commercially available product consists of polymers of two alternating disaccharide units, namely, D-glucosamine-D-glucuronic acid and D-glucosamine-L-iduronic acid. It is an anticoagulant that prolongs the clotting time of blood and inhibits the formation of fibrin clots both *in vivo* and *in vitro*. It exerts its activity by forming a complex with antithrombin III to accelerate the inactivation of thrombin and to inhibit other coagulation proteases such as factor Xa, which is responsible for the conversion of prothrombin to thrombin.

The biological assay for heparin sodium consists of comparing the activity of the heparin sample and USP Heparin Sodium Reference Standard in preventing the clotting of citrated sheep plasma. The USP unit of heparin is the concentration of heparin that will inhibit 1.0 mL of citrated sheep plasma from clotting up to 1 hour after the addition of 0.2 mL of $CaCl_2$ solution (1:100). Because the potency of heparin varies from different preparations, heparin always should be expressed and prescribed in units, rather than by weight.

Heparin Sodium Injection USP is used in the prophylaxis and treatment of venous thrombosis and pulmonary embolism, in atrial fibrillation with embolization, and for the prevention of clotting in cardiac and arterial surgery. Commercially available preparations of Heparin Sodium Injection are obtained from bovine lung and porcine intestinal mucosa. Heparin Lock Flush Solution is prepared from porcine intestinal mucosa and is a sterile preparation of Heparin Sodium Injection containing sodium chloride in an amount to make it isotonic with blood. This preparation of heparin is used for clearing intermittent infusion sets. Heparin Calcium and Heparin Calcium for Injection are also available and are obtained from porcine intestinal mucosa.

PROTAMINE SULFATE

Protamine Sulfate is a mixture of simple proteins of low molecular weight that are rich in arginine. They are found in the sperm or mature testes of salmon and various other species of fish. Due to the high content of arginine, the protamines are strongly basic. In the absence of heparin, the intravenous administration of protamine exerts an anticoagulant effect through its interaction with platelets and fibrinogen. In the presence of heparin, protamine and heparin interact to form a stable salt that results in the loss of anticoagulant activity of both drugs.

The biological assay for protamine sulfate depends on the ability of protamine to neutralize the anticoagulant activity of heparin in citrated sheep plasma. In this assay, various amounts of heparin are added to plasma containing a constant concentration of protamine. A solution of calcium chloride containing thromboplastin is added to the above samples and the clotting times are monitored. A detailed description of this assay can be found in the USP. Each milligram of protamine sulfate, calculated on the dried basis, neutralizes not less than 100 USP Units of heparin activity derived from lung tissue or intestinal mucosa. Protamine Sulfate for Injection is a sterile mixture of protamine sulfate with one or more suitable, dry diluents. Protamine Sulfate Injection, a sterile isotonic solution of Protamine Sulfate, is used in the treatment of heparin overdosage.

VITAMINS

Chemical or spectrometric assay procedures are specified for all preparations of vitamin A, vitamin B_1 (thiamine), and vitamin D. A biological method for assaying vitamin D has been recently approved by the USP (4th Supplement, USP 23-NF 18). This assay measures the ability of vitamin D to stimulate calcification of the rachitic metaphysis in rats. Because calcification depends on adequate amounts of vitamin D, rats fed a diet deficient in vitamin D develop rickets; supplementation of the diet with adequate amounts of vitamin D results in recalcification of the bone. Briefly, the assay uses young rats (not older than 55 days) who have developed rickets on a *rachitogenic diet*. These rats are divided into groups and fed the rachitogenic diet that has been supplemented with either USP Cholecalciferol Reference Standard, unknowns, or no supplementation (control).

One half of the dose of vitamin D either as the USP Cholecalciferol Reference Standard or unknown is given to the rats on day 1 and day 3 (or 4) of the assay period. At the end of a fixed period (between 7 and 10 days) the rats are weighed and sacrificed. Any rats whose weight has decreased are removed from further analysis. The leg bones of the remaining rats are dissected out and assayed for amount of recalcification of the bones. The activity of the vitamin D may be determined by the amount of recalcification relative to the reference standards.

Additionally, a biological assay for determination of vitamin D activity of Cod Liver Oil, Nondestearinated Cod Liver Oil, Oleovitamin A and D, and descriptions of formerly official biological assay methods for vitamins A and B_1 will be found in the 13th edition of this text (*RPS*-13, pp 1600–1604).

SUMMARY TABLE

Major aspects of the biological assay procedures for several official articles are summarized in Table 31-1.

MICROBIAL ASSAYS

As previously noted in this chapter, *biological assay* refers to measurement of the relative potency or activity of compounds by determining the amount required to produce a specific, defined effect on a suitable test animal or organ under standard conditions. The experimental animals mentioned in specific test procedures described in the previous section include mice, rats, guinea pigs, rabbits, cats, dogs, and pigeons. As noted earlier, a biological assay may involve observations or measurements of effects obtained in any form of living matter, plant or animal. The term *microbial* (a contraction of microbiological) *assay* designates a type of biological assay, specifically, a biological assay performed with *microorganisms,* such as bacteria, yeasts, and molds.

In general, the principles involved in microbial assays are the same as those that apply to assays using higher forms of plant or animal life. One notable difference involves the relative size of the experimental population. In the bioassays described above, the response of each individual test animal is noted and the results are obtained when a series of animals are subjected to statistical analysis to calculate mean activity, standard error, and so on. In a typical microbial assay, each evaluation is performed with a culture of microorganisms, and the measurement represents the average response of an extremely large population of test organisms. In the case of most bioassays, a linear relationship exists between the *log dose* and the response, whereas in most microbial assays there is a linear relationship between the *dose* and the response (within certain limits). The importance of this relationship in the evaluation of microbial assays is considered in Chapter 12.

VITAMINS

Microbiological procedures are available for the assay of Calcium Pantothenate, Dexpanthenol, Niacin or Niacinamide, and Vitamin B_{12} activity of Cyanocobalamin Co 57 Solution and Capsules, and Cyanocobalamin Co 60 Solution and Capsules.

A fundamental requirement in a microbial assay for the activity of a vitamin or amino acid (factor) is the inability of the test organism to synthesize the factor being assayed. Furthermore, the test organism must require the factor being assayed for normal growth, and should be sensitive to very small amounts of the required factor. For these types of microbial assays, special media are prepared that are nutritionally complete in all respects except for the factor under study. Examples of these media may be found in the USP section <81>. Control tubes containing the suitable media inoculated with the test species exhibit no, or only minimal, growth. If the basic requirements specified above are satisfied, the growth response of the test organism is, within limits, proportional to the amount of factor added to the medium.

The extent of the growth response may be determined by turbidity or spectrometric measurement, or by titration of the acid produced as metabolic waste. The turbidity of the culture is proportional to the amount of microbial growth; the development of acidity also reflects quantitatively the growth response.

Sufficient levels of reference standard are included to enable construction of a dose response curve for each assay. The activity of the factor (or factor dilution) being tested is determined by interpolation from the standard curve.

Niacin or Niacinamide

The techniques and procedures used in the microbiological assay for niacin are common to many of the microbiological methods, and a description of the niacin method will serve to give the pattern generally employed.

THE MICROORGANISM

In the case of niacin, it has been clearly demonstrated that the assay organisms used metabolize only the forms of niacin that are available to the host in which they normally grow. The fact that some organisms are more limited than the host animal in their ability to use niacin derivatives serves as a basis for differentiating such compounds in biological materials. For example, in addition to the free niacin, *Lactobacillus plantarum* is able to use niacinamide, nicotinuric acid, cozymase, and niacinamide nucleoside.

Although a number of microorganisms require niacin for their metabolic processes and are unable to synthesize it for themselves, the acid-forming organism *L plantarum* is used most widely for assay purposes. It is nonpathogenic, easy to culture, and affected to only a limited degree by stimulatory or inhibitory substances normally found in foods or pharmaceutical preparations containing niacin. It may be grown on a simple stab-culture medium containing gelatin, yeast extract, and glucose, and is cultured for use in the assay tubes by direct transfer to the liquid medium consisting of the basic assay medium containing an optimum amount of added niacin.

One important advantage of microbiological procedures is that only a minute quantity of a vitamin is needed to give a measurable response. For example, the range of niacin added to the series of standard tubes is 0.05 to 0.50 µg/tube. Thus, the niacin content of extremely small amounts of biological materials may be measured readily. Modifications using microanalytical apparatus and a lower range of vitamin additions have been described for blood and tissue analysis.

THE TEST SOLUTION

The first step in the assay procedure is the preparation of the test solution of the material to be assayed. If the sample is a dry or semisolid material, the niacin is extracted by heating the sample in a measured volume of dilute H_2SO_4 in an autoclave for 30 minutes. Liquid preparations are autoclaved 30 minutes after addition of the H_2SO_4 to give a concentration of 1 N H_2SO_4. Although niacin is soluble in water, certain precursors, found particularly in cereals, are unavailable to the test organ-

ism unless hydrolyzed. Either acid or alkali is equally effective for the extraction but acid is preferred, owing to the possibility of hydrolysis of trigonelline in an alkaline solution. Preparation of the test solution is completed by neutralizing with strong NaOH solution, then diluting to a volume that contains 0.1 µg of niacin per milliliter. Further purification of the test solution is not ordinarily important, because *L plantarum* is relatively unaffected by substances that inhibit or stimulate other test organisms.

THE MEDIUM

The basic medium employed in a niacin assay is simple to prepare and, with properly treated casein hydrolysate, is otherwise nutritionally complete. Both dehydrated complete media and dehydrated casein hydrolysates are available commercially and appear to be entirely satisfactory for assay purposes. To prepare a medium suitable for assaying amino acids, the casein hydrolysate is replaced with an amino acid mixture, omitting only the amino acid being assayed.

Details of the microbial assay procedure for niacin (including preparation of standard niacin solution, spectrometric determination of cell density and calculation of the niacin content of the test samples) are given in the official compendium.

Calcium Pantothenate

In the assay of Calcium Pantothenate, stock complete media supplemented with casein hydrolysate and additional vitamins, as described in the USP, are prepared. Sets of tubes are supplemented with Calcium Pantothenate standards or test solutions. Inoculum of *L plantarium* are prepared in broth from agar slants. One drop of *L plantarium*–containing broth is used to inoculate 10 mL of the broth for test purposes. The amount of growth of the *L plantarium* culture is determined by measuring light transmittance in a spectrophotometer following 16 to 24 hours of incubation. A dose-response curve is drawn by plotting the transmittance for each level of the standard solution against the amount of Calcium Pantothenate in the respective tubes. The amount of Calcium Pantothenate contained in the test solution is determined by proper interpolation of the observed values with the standard curve.

Dexpanthenol

This assay is used to determine the amount of Dexpanthenol as an ingredient of multi-vitamin preparations or for identifying the amount of dextrorotary panthenol in a racemic mixture of dextro- and levo- panthenol.

Briefly, a modified panthenate medium is made and a set of tubes is supplemented with stock Dexpanthenol or unknown, 0.5mL of a standard stock culture of *Pedicoccus acidilactici* is added to 10mL of culture media and incubated for 16 hours. The cultures are then killed by heating and their turbidity is determined spectrophotometrically. A dose-response curve is plotted and the concentration of the unknown is determined by interpolating the unknown values with the standard curve.

Vitamin B$_{12}$ Activity

Determination of vitamin B$_{12}$ activity requires special treatment of the material to be assayed so that the vitamin is made available to the test organism, which is a culture of *Lactobacillus leichmannii*. The basic medium used is quite complex, prepared as a mixture in solution of a variety of essential nutrients. Measured amounts of the material to be assayed are added to one set of tubes containing this medium, and measured amounts of the Standard Cyanocobalamin Solution are added to a corresponding second set of tubes. The tubes are inoculated with a small amount of culture of the test organism and then incubated overnight. The extent of growth of the microorganisms is measured by determining light transmittance by means of a spectrometer. A concentration-response curve is drawn as described for Calcium Pantothenate, and the amount of vitamin B$_{12}$ contained in the test solution is determined by proper interpolation of the observed values on the standard curve.

ANTIBIOTICS

The term *antibiotic,* as used in the official compendia, designates a medicinal preparation containing a significant quantity of a chemical substance that is produced naturally by a microorganism, or artificially by synthesis, and that has the capacity to inhibit or destroy microorganisms in dilute solution. Under the terms of the Federal Food, Drug, and Cosmetic Act of 1938, batch certification for antibiotics, whether for human or veterinary use, was introduced in stages: 1945, penicillin; 1948, streptomycin; 1949,aureomycin, bacitracin, and chloramphenicol. In 1962, as part of the Kefauver-Harris Amendments, batch certification was applied to all antibiotics intended for human use. On 1982 the FDA issued regulations that exempted antibiotics from batch certification requirements so long as the articles complied with standards; however, Section 507 (Certification of Antibiotics) remains intact.

Standards of potency and purity for antibiotics are established by the FDA in the form of regulations published from time to time in the *Federal Register*. Because all recognized antibiotics are subject to the provisions of the regulations, these determine the official standards. The federal regulations governing all aspects of antibiotic testing are extremely detailed and are subject to periodic amendment; they should be consulted with regard to prescribed methods for the assay of individual antibiotics and their preparations.

In evaluation of the potency of antibiotic substances, the measured effect is inhibition of the growth of a suitable strain of microorganisms—that is, the prevention of the multiplication of the test organisms. The procedures employed in microbial assay of antibiotics may be divided into two broad classifications: the *Cylinder-Plate Method* and the *Turbidimetric Method*.

Cylinder-Plate Method

The Cylinder-Plate Assay of antibiotic potency is based on measurement of the diameter of zones of microbial growth inhibition surrounding cylinders containing various dilutions of test compound, which are placed on the surface of a solid nutrient medium previously inoculated with a culture of a suitable organism. Inhibition produced by the test compound is compared with that produced by known concentrations of a Reference Standard.

Turbidimetric Method

The Turbidimetric Assay of antibiotic potency is based on inhibition of microbial growth as indicated by measurement of the turbidity (transmittance) of suspensions of a suitable microorganism in a fluid medium to which have been added graded amounts of the test compound. Changes in transmittance produced by the test compound are compared with those produced by known concentrations of reference material.

Detailed descriptions of appropriate microbial assays for specific antibiotics (ie, cylinder-plate or turbidometric method) may be found in section <81> of USP 25-NF 20. This section also catalogs the test organisms to be used with each antibiotic.

BIOLOGICAL TESTS

In the context of *bioassay,* a biological test has as its objective the qualitative determination of a specific characteristic of a biological product or of the container in which it is supplied (eg, transfusion assemblies). These tests are designed to determine with a high degree of certainty the absence or presence of a type of activity (such as antibacterial activity or pressor activity), or quality (such as nonantigenicity or toxicity), or constituent (such as depressor substances or pyrogens). Animals are employed in some tests and microorganisms in others.

PYROGEN TEST

The USP pyrogen test requires healthy, mature rabbits to determine the absence or presence of pyrogens in products that can be tolerated by the rabbit. Three rabbits are used; each receives 10 mL of the test solution/kg by injection into an ear vein, completing the injection within 10 min. The rectal temperature is recorded at 1, 2, and 3 hr after the injection. The Decision Statements specify a limit on the temperature rise allowed for any one rabbit as 0.5° If a single rabbit has an increase in temperature 0.5° or greater then the test is expanded to include five additional rabbits (for a total of eight animals); after which the requirement for absence of pyrogen states that no more than three rabbits each exhibit a temperature rise of less than 0.5°, and the total temperature rise for all eight rabbits is 3.3° or less.

BACTERIAL ENDOTOXIN TEST

Since USP 21-NF 15, an alternative to the rabbit pyrogen test exists in the form of the Bacterial Endotoxin Test (BET). In this *in vitro* procedure, the aqueous extract of the circulating amebocytes of the horseshoe crab, *Limulus polyphemus*, called Limulus Amebocyte Lysate (LAL) is used because it causes the formation of a gel-clot if pyrogen (bacterial endotoxin) is present above a limiting concentration. Briefly, an aqueous sample is mixed with the LAL and incubated at 37°. The end point of the assay is determined spectrometrically by an increased turbidity due to gel formation or by the presence of a clot. This procedure requires a USP Reference Standard (defined potency of 10,000 USP Endotoxin Units per vial, endotoxin obtained from *Escherichia coli*) and provides a more sensitive detection of pyrogen (endotoxin) than does the rabbit test. Another advantage of the BET is that the presence of pyrogens may be detected in drugs that have definite physiological effects and for which the classic rabbit response could not be used.

The BET procedure has been widely adopted; for example, 216 monographs that used the rabbit pyrogen test now require the BET procedure instead. In 164 cases where the rabbit pyrogen test was unreliable, BET requirements are now found. BET is an ideal assay for ensuring that water used for pharmaceutical purposes is pyrogen free.

A discussion of pyrogens may be found in Chapter 40. This section discusses the nature, sources, and means of destruction of pyrogens. In addition, the testing for pyrogens is considered.

DEPRESSOR SUBSTANCES TEST

In the Depressor Substances Test, a female nonpregnant adult cat is anesthetized and the carotid or other suitable artery is exposed, separated from surrounding tissue, and monitored for continuous blood pressure. A femoral artery is exposed as a means to facilitate the intravenous injections of standard and test drugs. In this procedure, the depressor responses of the substance under test is compared to those responses elicited by several doses of a Standard Solution of Histamine. The substance under test is dissolved in a designated diluent to give the required concentration specified in the individual monograph. Refer to Biological Tests in the USP for the specific experimental steps required by the Depressor Substances Test.

BIOLOGICAL REACTIVITY TESTS, *IN VITRO* AND *IN VIVO*

In order to find suitable *in vitro* replacements for animal procedures, in 1985 the United States Pharmacopeial Convention established a standing group, the Subcommittee on *in Vitro* Toxicity, that has made impressive progress in gathering information about and in stimulating research into *in vitro* methods. The work of the Subcommittee has resulted in two proposed chapters, Biological Reactivity Tests *in Vitro*, and Biological Reactivity Tests *in Vivo*. Biological reactivity is the response of a biological system or the products of a biological system, such as cells from tissue culture, to an imposed stimulus.

In the Biological Reactivity Tests *in Vitro*, three tests called the *Agar Diffusion Test*, the *Direct Contact Test,* and the *Elution Test* are used to evaluate the suitability of elastomers and other polymers that are to have direct or indirect patient contact. The choice of test is dependent on the material, final product, and intended use.

Agar Diffusion and Direct Contact Tests

The Agar Diffusion Test is designed to assay elastomeric closures. In this assay, an agar layer protects tissue culture cells from contact with the material but allows diffusion of leachable chemicals from the polymers to reach the cells. To perform the test, monolayers of cultured cells are grown to 80% confluence in 60-mm diameter plates and the tissue culture media is replaced by media containing not more than 2% agar. After the agar has solidified, filter paper to which extracts from the polymers have been applied are placed on the agar.

The Direct Contact Test is designed for materials with a variety of shapes and also uses 80% confluent tissue culture cells. In the Direct Contact Test, the monolayers are incubated in direct contact with the samples. For both the Agar Diffusion and the Direct Contact Tests the cells are incubated with sample preparations or USP Negative Control Plastic RS or USP Positive Bioreaction Solid RS. After 24 hours of culture, the monolayer is examined microscopically and the observations made according to prescribed guidelines. The response is defined in terms of a series of Reactivity Grades.

Elution Test

The Elution Test evaluates the effects of extracts of polymeric materials on cultured cells. The materials are extracted into cell culture media (with or without serum supplementation) at either physiologic or nonphysiologic temperatures. When the monolayers are 80% confluent, the media is replaced with media containing the polymer extracts and the cells are incubated for an additional 48 hours at 37°. The evaluation once again is based on the microscopic appearance of the cells after incubation.

Systemic Injection Test

The Systemic Injection Test uses mice that receive an intravenous (IV) or intraperitoneal (IP) injection of a defined quan-

tity of an extract of a plastic, after which the mice are observed just after the injection and then after 4, 24, 48, and 72 hours. The animal response from injection of the sample is compared with the animal response from injection of a blank (same quantity of the same extraction medium treated the same way). The sample passes the test only if the animals do not display biological reactivity significantly greater than that seen in the blank.

Intracutaneous Test

The Intracutaneous Test determines the local responses to the polymer extracts following intracutaneous injection into rabbits. The evaluation is based on erythema and eschar formation and on edema production and uses a defined response value to label the intensity of the biological reactivity. The difference between the sample and blank must be less than one for the test to be met.

Implantation Test

The Implantation Test is a procedure that evaluates the suitability of the polymeric materials intended for containers or container accessories for use in parenterals and for use in medical devices, implants, and other related systems that may come into direct contact with living tissue. This test uses adult rabbits into whose paravertebral musculature are implanted strips no smaller than 10×1 mm. The animals are maintained for at least 120 hours following implantation, after which they are sacrificed and the tissue around the implant examined grossly and microscopically for reactions to the implant. Reference standard material called USP Negative Control Plastic RS is provided to act as a control.

Safety Tests—Biologicals

The Safety Tests—Biologicals is a procedure used to determine the acceptability, in terms of safety, of biologicals and biotechnology derived products. A dose of 0.5 mL of a test solution containing the biological is injected IP into mice and guinea pigs, after which they are observed for 48 hours or 7 days, depending on the test.

Because biologicals derived from biotechnology may include contaminants from the cell lines as well as adventitious infectious agents, the Federal Government has drafted a number of *Points to Consider* for use in characterization of monoclonal antibodies and characterization of cell lines used to produce bio-

Table 31-2. Summary of Biological Test Procedures

COMPENDIAL ARTICLE	ACTIVITY ASSAYED	ANIMAL EMPLOYED	ROUTE OF ADMINISTRATION OF TEST MATERIAL	END POINT OF TEST PROCEDURE
Iron dextran injection	Absorption of iron compound	Rabbit	IM	No heavy black deposit of unabsorbed iron 7 days after injection
Diphtheria toxoid, tetanus toxoid and combinations with pertussis vaccine	Antigenicity	Guinea pig	SC	Not less than 80% survival (for at least 10 days) of immunized animals injected with test doses of toxin
Protein hydrolysate injection	Nutritional completeness	Rat	PO	Weight gain while maintained on test product and nitrogen-deficient diet
Isofluorophate ophthalmic ointment	Miotic	Rabbit	Ocular instillation	Pupil constriction
Insulin products	Lower blood sugar level	Rabbit	SC	Glucose analysis
Technetium	Tc99m–containing compounds	Distribution of radioactivity	Rats or mice IV	Residual radioactivity in specified tissues
Diphtheria toxoid, tetanus toxoid	Toxin poisoning	Guinea pig	SC	No symptoms of toxin poisoning within 21 days
Many articles	Pyrogen test or bacterial endotoxins test (BET)	Rabbits (Pyrogen)	IV (Pyrogen)	Rectal temperature increase not more than 0.6° (Pyrogen)
		LAL	*In vitro*	Increased turbidity by spectrophotometry or clot formation
Elastomeric closures, plastic containers, transfusion assemblies	Systemic toxicity of extract	Mouse	IP, IV	No toxic reaction within 72 hr
	Intracutaneous toxicity of extract	Rabbit	Intracutaneous	No significant irritation compared to blank
	Implantation toxicity of designated material	Rabbit	Aseptic implant	No significant encapsulation compared to blank
	Agar Diffusion, Direct Contact Elution Test	Tissue culture	*In vitro*	Minimal damage to monolayer cells
Biologics and - biotechnology derived products	Toxicity	Mice Guinea pigs	IV, IP	Abnormal or untoward toxicity or death

Note: These tests are described in detail in the USP/NF and the Official Supplements.

logicals. The current approach is to focus on the production, identification, and characterization of the cells being used to produce the biological, the validation of the manufacturing process, and the testing of bulk and final product for safety. With regard to safety, the *Points to Consider* recommend testing for bacteria, fungi, and mycoplasma as well as tests for adventitious viruses including lymphocytic choriomeningitis virus, Epstein-Barr virus, cytomegalovirus and hepatitis B and C. If the cell lines being used are nonhuman lines, the recommendation includes assaying for viruses appropriate to the cell line. In addition, samples should be tested for retroviruses. Should viral contamination be found, the manufacturer should include steps to inactivate or remove the virus and these steps should be validated.

Medical devices labeled nonpyrogenic that make contact directly or indirectly with the cardiovascular system or other soft body tissue must meet the specifications for sterility, nonpyrogenicity, and safety as outlined under *Transfusion and Infusion Assemblies* in the USP. The pyrogen procedure uses the Bacterial Endotoxin Test (BET) using Limulus Amoebocyte Lysate (LAL) and, when appropriate, the rabbit pyrogen test. The mouse safety tests referred to before are used for extracts of the plastic.

Plastics for use in parental preparations, elastomeric closures for injections, and plastic containers for ophthalmics are subject to both *in vitro* and *in vivo* Biological Reactivity Tests. Plastic material from the three categories is tested for *in vitro* biological reactivity in the Agar Diffusion Test, Direct Contact Test, and Elution Test. Materials from plastics for use in parenteral preparations and elastomeric closures for injections that meet the *in vitro* tests are not required to undergo *in vivo* testing. Materials in the plastics for use in parenteral preparations category that do not meet the *in vitro* test requirements are subjected to the Systemic Injection, Intracutaneous, and Implantation Tests. Materials in the elastomeric closures for injections category that do not meet the *in vitro* test requirements are subjected to the Systemic Injection, Intracutaneous, and Pyrogen Tests. Materials intended for use in plastic containers for ophthalmics that do not meet the requirements of

the *in vitro* tests are tested by the Systemic Injection and Intracutaneous Tests. Materials cannot be used for containers for ophthalmic preparations if they do not meet the requirements of the Systemic Injection and Intracutaneous Tests.

The USP search for alternatives to *in vivo* tests is an ongoing activity. Work is in progress for selective replacements for the Eye Irritation procedure: methods involving the irritation of the chorioallantoic membrane (from the chicken egg), the uptake of neutral red dye by living cell lyosomes, the total cell protein assay, and the rabbit corneal epithelial cell healing and other characteristics. Two bacterial tests are under study to determine their potential value, *v12,* the Bacterial Bioluminescence Test, a toxicity procedure applicable not only to extracts of plastics but also to solutions of bulk pharmaceuticals; and the bacterial colorimetric test, a procedure based on β- galactosidase measurement and with potential value in the determination of the biological reactivity of chemicals and extracts of plastics.

SUMMARY TABLE

An outline of compendial articles subject to identification, activity, or toxicity tests of a biological nature is presented in Table 31-2.

REFERENCES

1. *Guide for the Care and Use of Laboratory Animals.* NIH Publ No 86-23. Bethesda, MD: US Department of Health and Human Services, 1985.
2. Bangham DR, Mussett MV, Stack-Dunne MP. *Bull WHO* 1962; 27: 395.
3. Sayers MA, Sayers G, Woodbury LA. *Endocrinol* 1948; 42: 379.

GENERAL REFERENCE

USP 25/NF 20. Rockville, MD: USP Convention, 2002; and supplements.

Clinical Analysis

Cathy Y Poon, PharmD

The characterization and quantitation of the various components of blood, urine, and other body fluids are the primary functions of the clinical laboratory. The major divisions of clinical analysis are clinical biochemistry, hematology, blood-bank technology, histopathology, immunology, and microbiology. The accurate diagnosis of disease and determination of a potential therapeutic regimen frequently are based on the laboratory analysis of blood, urine, feces, gastric secretions, or cerebrospinal fluid. Modern medical practice is tending toward greater reliance on laboratory results as definitive measures of pathological or normal states.

Pharmacists should familiarize themselves with the basic principles involved in sample collection, analysis, and diagnostic significance of the various clinical parameters. Their role in community health necessitates comprehension of the methodology and diagnostic value of clinical laboratory procedures. The influence of various drugs and drug interactions on these parameters must be considered in both the clinical and drug-abuse situation.

HEMATOLOGY

The determination of the morphological, physiological, and biochemical properties of peripheral blood and the blood-forming organs (hematopoietic system) is a function of the hematology laboratory. The functional categories of hematology are (1) analysis of cellular elements, and specific biochemical and physiological parameters of peripheral blood and the hematopoietic system, (2) blood-coagulation analysis, and (3) blood-bank technology.

Peripheral blood is a biphasic liquid tissue system of cellular elements suspended in a liquid plasma phase. The cellular phase comprises about 45% of the blood volume and contains erythrocytes (red blood cells, RBC), leukocytes (white blood cells, WBC), and thrombocytes (platelets). The plasma phase is primarily water (90–92%) and protein (7%).

The hematological analysis of blood is concerned primarily with enumeration and differentiation of the various cellular elements. An analysis of the hematopoietic system (eg, bone marrow and lymphoid tissue) determines the status of blood-cell precursors in these tissues. Determinations of specific biochemical (hemoglobin) and physiological (blood or plasma volume) parameters are performed in a complete evaluation of the erythron system (blood and marrow RBC and their precursors). The normal hematological values in the adult are presented in Table 32-1.[1]

ERYTHROCYTES AND HEMOGLOBIN

The erythrocytic system is composed of the mature erythrocytes in peripheral blood and their precursors in bone marrow. The precursors of erythrocytes, as found in the erythropoietic system (red bone marrow), are classified as to the degree of nucleation and characteristics of cytoplasmic constituents. The sequence of erythrocyte formation in bone marrow—based on the gradual denucleation of the cell, generation of the chromatin structure, and changes in nucleolar structure and cytoplasmic constituents—is

pronormoblast → basophilic normoblast →
polychromatic normoblast → orthochromatic normoblast →
polychromatophilic erythrocyte → erythrocyte

The first four types are nucleated and normally are seen only in bone marrow. In normal erythrocyte formation these immature bone-marrow cells are designated as *normoblastic* or *normocytic*. In pernicious anemia and related conditions, they become abnormally large and are designated *megaloblastic* or *megalocytic*. In iron-deficiency anemia, these cells become abnormally small and are designated *microblastic* or *microcytic,* of the iron-deficiency type.

Normal blood contains 0.5% to 1.5% of circulating erythrocytes as reticulocytes. These cells contain a fine network of basophilic reticulum that is demonstrable on staining with a vital dye such as brilliant cresyl blue. The number of these cells in the blood is a measure of effective erythropoiesis. High-circulating reticulocyte values are an index of erythropoietic activity and are found in the first few days of life, after hemorrhage and after treatment of iron-deficiency or vitamin B_{12}-deficiency anemias.

The normal *erythrocyte* (normocyte) is a flexible, elastic, biconcave, enucleated structure with a mean diameter of 7.3 μm and a thickness near 2.2 μm. The chemical constituents of the red blood cell include water (63%), lipids (0.5%), glucose (0.8%), minerals (0.7%), nonhemoglobin protein (0.9%), methemoglobin (0.5%), and hemoglobin (33.6%). The primary function of the erythrocyte is transport of oxygen and carbon dioxide. The red cell membrane, a dynamic, semipermeable component of the cell, is associated with energy metabolism in the maintenance of the permeability characteristics of the cell to various cations (Na^+, K^+) and anions (Cl^-, HCO_3^-). The stroma of insoluble material that remains after red-cell disruption (hemolysis) constitutes 2% to 5% of the wet-cell weight; it is primarily protein (40–60%) and lipid (10–12%). The membrane includes stromatin (a fibrous or structural protein) and mucopolysaccharides associated with A, B, and O blood-group substances. The lipid fractions include phosphatides (lecithin, cephalin), cholesterol, cholesterol esters, neutral fats, cerebrosides, and sialic acid glycoproteins.

Erythrocytes may be enumerated by either visual or electronic procedures. In the visual procedures, a measured quantity of blood is diluted with a fluid, which is isotonic with blood and will prevent its coagulation. The diluted blood is then placed in a counting chamber (hemocytometer), and the num-
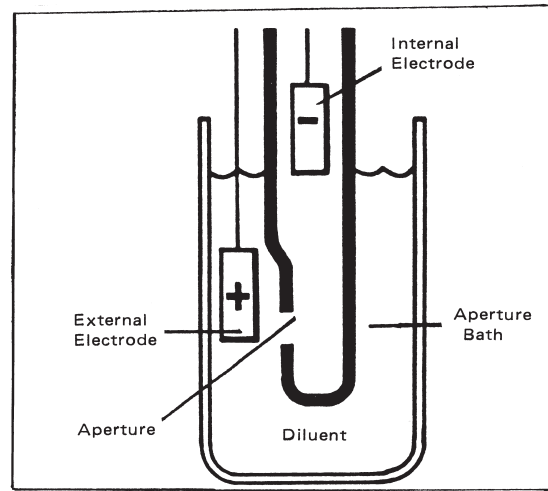
Table 32-1. Normal Hematological Values in Man

	NORMAL VALUE	NORMAL RANGE OF VALUES
Erythrocytes ($10^6/\mu L$)		
Male	5.2	4.2–6.1
Female	4.6	3.7–5.5
Reticulocytes ($10^3/\mu L$)	50	25–75
Hemoglobin (g/dL)		
Male	15.6	13.0–18.2
Female	13.6	11.0–16.3
Hematocrit (%)		
Male	45.0	36.5–52.0
Female	40.0	33.0–47.0
Mean corpuscular volume (fL)	88	75–100
Mean corpuscular hemoglobin (pg)	30	27–35
Mean corpuscular hemoglobin concentration (%)	34	31–37
Leukocytes ($10^3/\mu L$)	7.0	3.9–10.9
Leukocyte differential (%)		
Neutrophils	58	50–75
Bands	4	2–6
Eosinophils	2	1–5
Basophils	1	0–2
Lymphocytes	30	20–40
Monocytes	5	8–38
Platelets ($10^3/\mu L$)	300	150–450
Erythrocyte sedimentation rate (Westergren) (mm in 1 hr)		
Male	4	0–10
Female	10	0–20

From Simmons A. *Hematology—A Combined Theoretical and Technical Approach.* Philadelphia: WB Saunders, 1989, with permission from Elsevier.

Figure 32-1. Coulter-counting cells by electronic impedance (courtesy, Beckman Coulter).

ber of cells in a circumscribed area is enumerated microscopically. Hayem's solution (sodium sulfate, 2.5 g; sodium chloride, 0.25 g; mercuric chloride, 0.25 g; distilled water, 100 mL), Toison's fluid (sodium sulfate, 8 g; sodium chloride, 1 g; methyl violet, 0.025 g; glycerin, 30 mL; distilled water, 180 mL) or 0.9% sodium chloride are used as diluting fluids. The overall error of this method is about 8%.

A greater degree of accuracy and reproducibility can be achieved by erythrocyte enumeration in an electronic counting apparatus, such as the Coulter Counter or various flow cytometric instruments. The Coulter method (Fig 32-1) determines the number and size of particles suspended in an electrically conductive liquid. The blood cells traverse a small aperture and displace their own volume in the diluent as to produce a change in resistance between the electrodes; the magnitude of the voltage pulse is proportional to cell volume, and the resultant pulses are then amplified, scaled, and automatically counted.

In instruments such as the Bayer ADVIA 120 (Fig 32-2), the principles of laser flow cytometry are used to count cells. Hydrodynamic focusing and laminar flow are combined in the system to count a large number of individual cells. Light focused by a laser diode is scattered by the cells as they pass through the flow channel. The scattered light is monitored by a photoelectric sensor and transfers the electrical pulses, which are processed by the systems circuitry. In addition to increased counting speed, the overall error of the electronic procedures is reduced to about 1%.

The *hematocrit value* is also a measure of the erythrocyte portion of blood. A sample of blood containing an anticoagulant
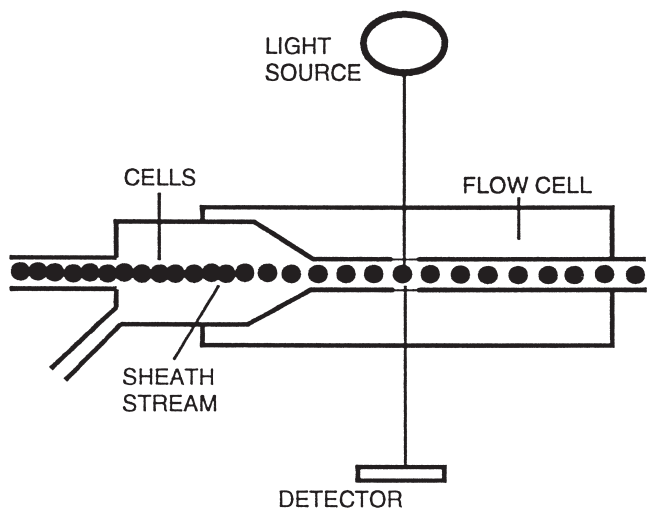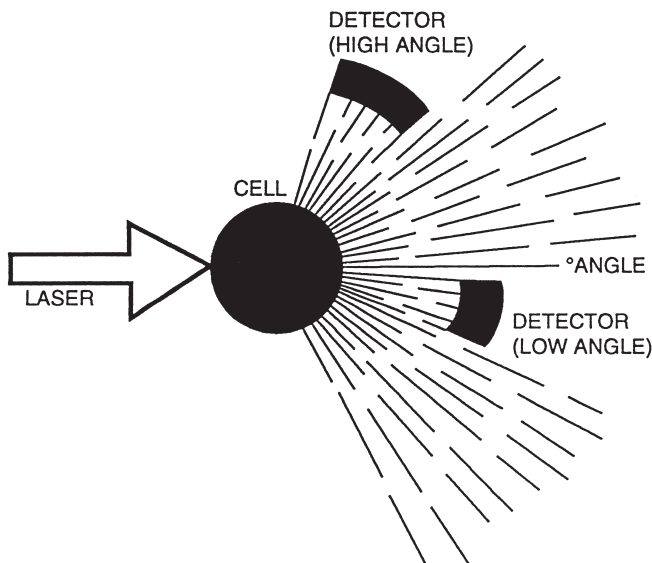
Figure 32-2. Sheath stream flow cell accomplishes hydrodynamic focusing, where cells pass single file through the detection area. Sensors detect high- and low-angle laser light-scatterer (courtesy, Bayer).

is placed in a graduated hematocrit capillary tube, centrifuged, and the volume ratio of packed red cells to total blood volume (hematocrit value) determined. The centrifuged sample appears as a red layer of packed erythrocytes over which is found an off-white layer of packed leukocytes and platelets, and a supernatant plasma phase. The hematocrit value is an index of both the number and size of the red cells.

Hemoglobin, a conjugated hemoprotein with an approximate molecular weight of 67,000 daltons, contains basic proteins, the globins, and ferroprotoporphyrin (heme). It is essentially a tetramer, consisting of four peptide chains, to each of which is bound a heme group. Heme, which constitutes about 4% of the weight of the molecule, consists of a divalent iron atom in the center of a pyrrole-porphyrin structure. Four distinct polypeptide chains (α, β, γ, δ) can be incorporated into hemoglobin. Normal adult hemoglobin is HbA = $\alpha_2^A\beta_2^A$. Fetal hemoglobin contains 2α and 2γ chains and is designated HbF = $\alpha_2^A\gamma_2^F$.

Differences in the structural sequences of amino acids in the peptide portion of the hemoglobin molecules are controlled genetically and are responsible for different types of hemoglobin. Based on the characteristic mobility of the hemoglobin, in an electric field (electrophoresis) on starch, paper, cellulose acetate, agar, or acrylamide gel media, many hemoglobin types have been recognized (see Chapter 33). Only types P, F, and A_1–A_4 are considered normal. Sickle-cell anemia and β- thalassemia are hemolytic anemias associated with abnormal hemoglobins (ie, Type S in sickle-cell anemia and abnormal production of the β chain in β-thalassemia). In homozygous *HbS disease,* sickling of the red cells is due to the low solubility of the abnormal hemoglobin in its reduced state, with the production of semicrystalline bodies (tactoids), which distort and elongate the cells. In the sickle-cell trait (heterozygous), the blood smear shows no sickle cells. In the homozygous condition, HbS accounts for nearly all of the hemoglobin with small amounts of HbF. In the heterozygous condition, HbS constitutes 50% or less of the hemoglobin, with the balance as HbA.

The detection of sickle-cell disease is performed by microscopic observation of the induction of red-cell sickling in the presence of a reducing agent such as sodium metabisulfite, or by quantitative determination of urea-dispersible turbidity induced by dithionite following reduction of HbS to deoxy-HbS in RBC lysates. The microscopic procedure will detect only homozygotes, whereas HbAS and HbS and its structural variant HbC-Harlem both are detected in the urea-dithionite technique. Commercial qualitative test kits are available for detecting sickle-cell trait and anemia by solubility determinations. All hemoglobins positive to the dithionite test must be electrophorized (cellulose acetate, citrate agar, or starch gel) to differentiate HbS from HbC and thalassemia traits. Drugs causing hemolysis in glucose 6-phosphate dehydrogenase (G6PD) deficiency include sulfones, nitrofurans, chloroquine, dimercaprol, nalidixic acid, and probenecid.

The *hemoglobin* concentration is measured spectrophotometrically after lysis of whole blood and conversion of hemoglobin to hematin, oxyhemoglobin, or cyanmethemoglobin. The addition of a strong base (NaOH) to pH 10 converts oxyhemoglobin, carboxyhemoglobin, and methemoglobin to hematin, which can be estimated photometrically. Weaker bases (Na_2CO_3 or NH_4OH) convert hemoglobin to oxyhemoglobin for analysis.

Total hemoglobin is measured also by conversion to cyanmethemoglobin using alkaline sodium cyanide–potassium ferricyanide reagent. Hemoglobin standards certified by the Clinical Standards Committee of the College of American Pathologists are used in these procedures, and all results are expressed as "grams hemoglobin per 100 mL (g/dL) blood."

In the normal state, the oxygen consumption of the RBC is low, and it is involved in the conversion of hemoglobin to oxidized (Fe^{3+}) methemoglobin (HbM), which cannot bind oxygen. The normal balance of HbM ($<0.5\%$) is maintained by two en-

zyme systems: NADH and NADPH methemoglobin reductases. An inherited deficiency of the RBC enzyme, G6PD, will increase the rate of reduction of glutathione and methemoglobin, make the cell more vulnerable to oxidative attack, and result in susceptibility to drug-induced or immune-mediated nonspherocytic hemolytic anemia. G6PD deficiency is found predominantly in persons of Mediterranean descent, Southeast Asians, Africans, and American blacks. The enzyme can be quantitated spectrometrically or by fluoronephelometry by measuring the rate of reduction of nicotinamide adenine dinucleotide phosphate (NADP) in the presence of G6PD. Presumptive screening tests based on reduced glutathione (GSH) content of blood before and after incubation with acetylphenylhydrazine also are used.

Erythrocyte count, hemoglobin content, and hematocrit value are used to determine various blood indices in the diagnosis and treatment of anemia. These measurements are

$$Mean\ corpuscular\ volume\{MCV(fL)\} = \frac{Hematocrit\ (\%) \times 10}{Erythrocyte\ count\ (10^6/\mu L)}$$

$$Mean\ corpuscular\ hemoglobin\ \{MCH\ (pg)\}$$
$$= \frac{Hemoglobin\ (g/dL) \times 10}{Erythrocyte\ count\ (10^6/\mu L)}$$

$$Mean\ corpuscular\ hemoglobin\ concentration\ \{MCHC\ (\%)\}$$
$$= \frac{Hemoglobin\ (g/dL) \times 100}{Hematocrit\ (\%)}$$

Other parameters used to characterize red-cell variation include the red-cell distribution width (RDW). The RDW is calculated directly by the standard deviation and coefficient of variation from a red-cell size-distribution histogram. The difference in cell size may be used to monitor patients with pernicious or hemorrhagic anemia.

Anemias are classified as to red-cell volume and hemoglobin concentration. *Macrocytic* (large cell: MCV > 94), *normocytic* (normal cell: MCV, 82 to 92), or *microcytic* (small cell: MCV < 80) are the classifications according to cell volume. Cellular hemoglobin concentration categorizes the cells as to *hyperchromic* (MCHC > 38), *normochromic* (MCHC = 32 to 36), or *hypochromic* (MCHC < 30). Examples of anemias are

I. Hypochromic Microcytic—erythroid normoblastic anemia in bone marrow
   A. Iron Deficiency—low hemoglobin (Hbg) and RBC, low serum iron, high total iron binding capacity, absent hemosiderin
      1. Dietary—low iron intake
      2. Intestinal problems—decreased iron absorption
      3. Pregnancy, infants—increased iron requirements
      4. Iron loss—due to chronic hemorrhage, parasitic infections, GI tract lesions, excess menstrual bleeding
   B. Hereditary Sideroblastic—defect in the heme synthesis, an inability to utilize ingested iron
   C. Thalassemia—genetic abnormality that produces normal to increased HbgF and/or HbgA₂
II. Normochromic Normocytic
   A. Hemolytic—increased destruction of erythrocytes
      1. Autoimmune hemolytic
      2. Cold agglutinin hemolytic
      3. Mechanical destruction of RBCs
      4. Paroxysmal nocturnal hemoglobinuria
      5. Lymphomas and Hodgkin's disease
      6. Infections
   B. Hemoglobinopathies—abnormalities in structure of alpha or beta chains of hemoglobin molecule; normoblastic erythroid hyperplasia in bone marrow
      1. Sickle-cell
      2. Hemolysis
      3. Hemoglobin CC
   C. Acute Hemorrhage
   D. Other
      1. Aplastic Anemia, Leukemia, Malignancy
      2. Renal failure and drug-related anemias caused by chloramphenicol and antineoplastic drugs

III. Normochromic Macrocytic—due to deficiency of vitamin B_{12} or folate; bone marrow is hypercellular with increased erythroid precursors
1. Pernicious
2. Sideroblastic
3. Sprue—total iron-binding capacity is decreased; hemosiderin is increased in the bone marrow
4. Pregnancy

Determinations of the suspension stability of whole blood and erythrocyte fragility are useful adjuncts in the diagnosis of various diseases.

The *erythrocyte sedimentation rate* (ESR) is an estimate of the suspension stability of red blood cells in plasma; it is related to the number and size of the red cells and to the relative concentration of plasma proteins, especially fibrinogen and the α- and β-globulins. This test is performed by determining the rate of sedimentation of blood cells in a standard tube. Normal blood ESR is 0 to 15 mm/hr. Increases are an indication of active but obscure disease processes such as tuberculosis and ankylosing spondylitis. ESR is affected by anemia and does not respond linearly with changes in asymmetrical macromolecules such as fibrinogen and globins.

The *zeta sedimentation ratio* (ZSR) technique overcomes these disadvantages. It is based on a measure of the closeness with which RBC will approach each other after standardized cycles of dispersion and compaction.

The *erythrocyte fragility test* is based on resistance of cells to hemolysis in decreasing concentrations of hypotonic saline.

Increased osmotic fragility of the red cells is associated with various types of spherocytosis and acquired hemolytic anemia; increased resistance has been observed in thalassemia, sickle-cell anemia, and hypochromic anemia. The test can be performed manually by colorimetric estimation of hemoglobin released by hypotonic cell rupture or automatically in an instrument, which continually records the increase in light transmittance through a suspension of red cells in a continuously decreasing salt gradient during dialysis.

LEUKOCYTES

Mature *leukocytes* (white blood cells, WBC) in peripheral blood and their precursors in bone and lymphoid tissue comprise the leukocytic system. Various types of leukocytes are found in normal blood. Differentiation of the lymphocytic, monocytic, and granulocytic leukocyte types is based on cell size, color, chromatin structure, and cytoplasm constituents.

The primary function of leukocytes is the development of the various defense mechanisms and repair processes in inflammatory and immune-response mechanisms. The migration of leukocytes to the site of inflammation is associated with the release or activation of various biochemical substances (5-hydroxytryptamine, histamine, complement, immunoglobulins, prostaglandins, lysosomal enzymes). The tissue histiocyte or monocyte (macrophage) also can engulf and destroy foreign particles by the process of endocytosis and certain leukocyte types by phagocytosis.

The chemical composition of the leukocyte includes water (82%), nucleoprotein, phospholipids, and trace minerals. Enzyme content, glycogen, and histamine levels vary in the different types of white cells. Deficiency in enzymes associated with glycolytic metabolism (hexokinase) and increases in phosphomonoester hydrolases (alkaline phosphatase) have been observed in leukocytes of certain leukemia patients.

The precursors of granulocytic leukocytes are found in bone marrow and are classified according to the degree of cytoplasmic granulation, dye-affinity of the granules, and shape of the nucleus (Schilling, Arneth, or Cooke-Ponder Classification). As undifferentiated cells (myeloblasts) mature

promyelocyte → myelocyte → metamyelocyte →

band leukocyte → segmented leukocyte

metachromatic granules appear in the cytoplasm (granulocytes). All segmented leukocytes are motile, a requirement for participation in the inflammatory or phagocytic processes.

In the mature *basophilic* and *eosinophilic leukocytes*, these granules develop an affinity for a basic or acidic dye, respectively; those cells containing granules that do not stain are called *neutrophils*. In peripheral blood, the mature granulocytic cells are designated *polymorphonuclear leukocytes: neutrophilic, eosinophilic,* or *basophilic*.

The other types of white cells normally observed in peripheral blood have no granules and are classified as to size and shape into the *monocyte* and *lymphocyte*, which are formed in lymphoid tissue. The small lymphocyte is thymic-derived and is found in the circulation and germinal centers of lymphoid tissue. The origin of the large lymphocyte is a gut-associated lymphoid stem cell that can further differentiate into the immunoglobulin-producing plasmacyte. The interaction of thymic (T) and bone-marrow (B) lymphocytes is the basis for the development and maintenance of humoral and cellular immune mechanisms.

Leukocytes are enumerated by procedures similar to those used for erythrocytes. In the visual procedures, the blood is diluted with a fluid (3% *v/v* acetic acid) that lyses the red cells, and the total leukocyte count is determined microscopically. Eosinophils also may be analyzed differentially with a diluting fluid that renders the red cells nonrefractile and invisible, and lyses the base-labile leukocytes, leaving the base-stable eosinophils intact. A suitable diluting fluid for this purpose is Pilot's Fluid (propylene glycol, 50 mL; distilled water, 40 mL; 1% phloxine, 10 mL; 10% sodium carbonate, 1 mL; and heparin sodium, 100 units). Electronic counting procedures are similar to those for erythrocytes with the added advantages of speed, accuracy, and reproducibility.

The normal adult leukocyte value is 5000 to 10,000 cells μL^3. Values greater than 10,000 (*leukocytosis*) are encountered in the newborn infant, young children, leukemia, cancer, convulsive seizures of epilepsy, and after extreme exercise. Values of less than 5000 (*leukopenia*) are observed in certain microbial infections (eg, typhoid fever, measles, malaria, overwhelming septicemia), cirrhosis of the liver, pernicious anemia, radiation injury, and replacement of marrow by malignant tissue.

A *differential count of the leukocytes* provides information as to the relative numbers of each type. A thin film of blood is prepared on a microscope slide, stained with a polychromatic preparation such as the Leishman, Wright, or Giemsa stain, and analyzed microscopically. Wright's stain contains polychromed methylene blue and eosin dyes; the erythrocytes are stained pink; the nuclei of the leukocytes, purplish-blue; neutrophilic granules, violet-pink; eosinophilic granules, red; basophilic granules, blue; and platelets, blue.

The introduction of automated systems for differential white-cell counts has reduced the errors inherent with the subjective nature of the visual counting procedure. Differentiation of the various cell types can be made on the basis of cytochemistry and staining properties of enzymes specific for a single cell type. The granules of neutrophils and eosinophils are stained by action of their peroxidases on 4-chloro-1-naphthol to form a colored quinone in the presence of a peroxide and further differentiated by the optimum pH for peroxidase activity between these two cell types. The monocytic lipase is used as a specific marker by the reaction of basic fuchsin with α-naphthol liberated by lipase on α-naphthylbutyrate substrate. The lymphocytes are not stained in this procedure but are measured by electronic sizing.

Automated differential WBC counts also are obtained in systems that count large populations of cells by simultaneous measurement of two optical properties (axial light loss and/or narrow-angle scatter and/or multiple-wavelength fluorescence). Laser light also is used to differentiate cell size, granularity, and volume of cells. The collected light measured by forward versus right-angle scatter is converted to a histogram giving the percent of lymphocytes, monocytes, and granulocytes. Another type of system utilizes computer processing of two-dimensional images of the various cell types after staining, employing an automatic scanning microscope.

Polymorphonuclear neutrophilic leukocytes (neutrophils, "polys") normally comprise 62% (50–67%) of the total leukocyte count. These cells are irregular in shape (10–15 μm in diameter) and usually contain a multilobated nucleus with fine,

lightly stained cytoplasmic granules. An immature or juvenile form of neutrophil, with a band-shaped nonsegmented nucleus constitutes 3% to 5% of peripheral blood leukocytes. Increases in the relative percentage of these cells (neutrophilia) are observed in acute microbial infections (eg, meningitis, smallpox, poliomyelitis), metabolic disorders (diabetic acidosis, gout), drug intoxication (digitalis, epinephrine), vaccination, coronary thrombosis, and malignant neoplasms.[2]

Polymorphonuclear eosinophilic leukocytes (eosinophils) normally comprise about 1% to 3% of total circulating white blood cells. In appearance they are similar to the neutrophil with the exception of large, red-stained cytoplasmic granules. Eosinophilia has been observed in certain skin diseases (psoriasis, eczema), parasitic infestations (pork round worm—trichinosis), certain hypersensitivity reactions, and in scarlet fever and pernicious anemia. Charcot-Leyden crystals, which are found in bronchial secretions from asthmatics, are derived from nucleoprotein-disintegration products of eosinophils.

Polymorphonuclear basophilic leukocytes (basophils) possess large cytoplasmic granules that stain a deep blue. These cells, which are primarily sources of blood heparin and histamine, constitute less than 1.0% of the leukocytes. Basophilic leukocytosis is seen in chronic myelocytic leukemia, hemolytic anemia, and Hodgkin's disease. Basophilic leukopenia occurs following radiation or therapy with glucocorticoids.

Lymphocytes have a cell diameter from 7 to 10 μm (small) to 10 to 18 μm (large). They have a round, or slightly indented, deeply stained nucleus and normally comprise 25% to 33% of the leukocytes. Lymphocytosis is seen in infectious mononucleosis, lymphocytic leukemia, rickets, and in most conditions associated with neutrophilic leukopenia (neutropenia).

Monocytes constitute 3% to 7% of the leukocytes. They are larger (12–20 μm) than the other leukocytes and possess an abundant, pale, bluish-violet-stained cytoplasm with a fine, reticulated chromatin structure in the nucleus. The monocytes (macrophages) phagocytize bacteria, parasitic protozoa, foreign particles, and even erythrocytes. Monocytosis is seen in certain microbial infections (tuberculosis, typhus, malaria), Hodgkin's disease and monocytic leukemia.

Drug therapy frequently causes neutrophil dysfunction, which can be characterized by a decreased number of mature neutrophils or a defect in cellular function resulting in the inability of the body to defend itself against infection. Drugs such as nitrogen mustard and chloramphenicol degenerate bone-marrow stem cells, and DNA synthesis is impaired by antimetabolites such as methotrexate and fluorouracil. Depolymerization of DNA is caused by procarbazine and alkylating agents. Mitosis is inhibited by colchicine and vinca alkaloids. The following outline lists drugs that cause granulocytopenia.[2]

Nonchemotherapeutic
  rifampin
  ristocetin
  benzene
  nitrous oxide
  ethanol
Antithyroid
  carbimazole
  methimazole
  thiouracil
Diuretics
  acetazolamide
  chlorthalidone
  chlorothiazide
  ethacrynic acid
  hydrochlorothiazide
  mercurials
Antihistamines
  thenyldiamine
  thenalidine
  pyribenzamine

Phenothiazines
  chlorpromazine
  mepazine
  methotrimeprazine
  prochlorperazine
  thioridizine
Antibiotics
  chloramphenicol
  carbenicillin
  griseofulvin
  isoniazid
  novobiocin
Cardiovascular
  diazoxide
  procainamide
  methyl dopa
  quinidine
  propranolol

As qualitative and quantitative changes in leukocytes in peripheral blood and their precursors in bone marrow and lymphatic tissue are associated with the various types of *leukemia*, this disease has been classified on the basis of the predominating type of leukocyte, ie, myelocytic (granulocytic), lymphocytic, monocytic, or plasmacytic. Leukemia may be either acute or chronic and involve the replacement of bone-marrow elements by malignant cells, infiltration of the reticuloendothelial system, anemia, thrombocytopenia, and hemorrhage. Leukemia usually is associated with an elevated WBC count and increase in the specific cell and its precursors in peripheral blood, but in certain instances there is an aleukemic blood picture with no evidence of leukocytosis. Leukocytes in acute leukemia are more immature ("blast"-type cells) than those encountered in the chronic type.

In many diseases of the hematopoietic system, it is necessary to examine the bone marrow to determine the rates of formation, maturation, and release of blood cells into the peripheral circulation. Using a puncture biopsy needle, samples of *bone marrow* may be obtained from the sternum, iliac crest, or proximal end of the tibia. Smears of marrow then are prepared, stained (Wright's stain or specialized histopathological procedure) and examined microscopically. The ratio of myeloid leukocyte to nucleated red cells in bone marrow, the presence of abnormal (*nonmyeloid*) cells, the number of platelet precursors (*megakaryocytes*), the signs of cell-maturation arrest, and the presence of focal lesions are important factors in the diagnosis of various disease states.

Systemic lupus erythematosus (SLE) is a disease characterized by numerous clinical and pathological manifestations associated with various organs. Although the disease chiefly affects the lymphatic system, the cardiac, renal, and articular systems also are involved. The diagnosis of this disease is based on the presence of an SLE-cell factor in the gamma-globulin fraction of blood in the diseased state. This factor dissolves the nuclei of leukocytes by depolymerization of deoxyribonucleic acid to form the SLE-body. If serum from patients with SLE is incubated with white cells, the "polys" will engulf the liberated SLE-body and form the typical SLE-cell with a characteristic progressive loss of nuclear detail. Drugs that cause SLE and produce a positive SLE-prep include hydralazine, procainamide, isoniazid, and phenytoin.

These antibodies to nucleoprotein also can be detected by immunologic techniques. In the double-antibody technique, the test serum containing antibodies to nuclear protein is incubated with a rat kidney slice (antigen). The second antibody is a fluorescein-labeled goat antihuman immunoglobulin (IgG) that binds to the human IgG, which is also bound to the antigen site in a positive test. The fluorescence is estimated by immunomicroscopy. Normal light-microscopy can be used if the goat-antihuman IgG is labeled with peroxidase.

THROMBOCYTES

The primary functions of *thrombocytes* (blood platelets) are the maintenance of hemostasis (arrest of blood flow from a vessel) and blood coagulation (clot formation). Platelets are oval to spherical in shape and have a mean diameter of 2 to 4 μm. They originate from an immature cell (megakaryocyte) in bone marrow; ranges of 140,000 to 450,000 μL have been reported in normal blood.

Adhesiveness, aggregation, and agglutination are the principal physical properties of platelets responsible for hemostasis and coagulation reactions. Chemically, they contain protein (60%), lipid (15%), and carbohydrate (8.5%). Their content of serotonin, epinephrine, and norepinephrine aids in promoting constriction at the site of injury. The release of "platelet thromboplastin," a cephalin-type phosphatide, and adenosine diphosphate (ADP) are important in blood coagulation.

Manual methods for the enumeration of blood platelets are notoriously imprecise due to the size and physical properties of the platelet. Indirect methods of analysis are based on the proportion of platelets to erythrocytes in a stained blood smear. Blood samples obtained directly from the fingertip puncture are

diluted with an anticoagulant fluid that simultaneously will stain the platelets. The ratio of platelets to red cells then is determined microscopically, and the number calculated from the predetermined red-cell count (normal 3 to 8 platelets/100 RBC). In the direct procedures, a sample of blood is obtained by venipuncture, placed in a siliconized tube, diluted, and subsequently analyzed by counting the platelets in a microscopic counting chamber using conventional or phase-microscopy apparatus. Suitable diluting fluids are the Rees-Ecker Fluid (sodium citrate, 3.8 g; formaldehyde, 0.22 mL; brilliant cresyl blue, 0.05 g; water, qs 100 mL) or Brecker Fluid (1% ammonium oxalate). Automated procedures for platelet counting have increased the accuracy to ± 5% to 10%. Blood is collected in a special anticoagulant, diluted, and centrifuged at specified speeds to obtain a "platelet-rich" supernatant fluid, which then is counted in an automated counting apparatus similar to those used for RBC counting.

Methods for counting platelets in whole blood include electronic impedance instruments and laser-optical counters using hydrodynamic focusing.[3] These multiparameter hematology analyzers provide greater accuracy, precision, and increased rate of analysis performed on a small volume of blood. The automated instruments provide precise platelet measurements for monitoring chemotherapy-induced thrombocytopenia and transfusion therapy.

Persistent increases in platelet count (*thrombocythemia* or *piastrinemia*) have been observed in chronic myelocytic leukemia, polycythemia, megakaryocytic hyperplasia, and splenic atrophy. Acute or temporary increases in platelet values (*thrombocytosis*) are seen in trauma and asphyxiation.

Thrombocytopenia or a decrease in platelets to values less than 60,000/μL occurs in various purpuras or hemorrhagic states (idiopathic or symptomatic thrombocytopenic purpura). Inherited platelet defects include Glanzmann's thrombasthenia, which is characterized by prolonged bleeding time and poor clot retraction, whereas Bernard-Soulier Syndrome and Von Willebrand's disease demonstrates defective platelet adhesiveness. Defects in the release reaction include "Storage Pool Deficiency" and "Aspirin-like Syndrome."

A rare, inherited, structural, and functional platelet abnormality is the *grey-platelet syndrome,* characterized by large platelets lacking alpha granules and appearing grey on Wright's-stained peripheral blood smears. Patients have a history of bleeding, petechiae, easy bruising, and epistaxis. Diagnosis is confirmed by radioimmunoassay procedures to detect levels of platelet-specific alpha-granule proteins.

Leukemia, extensive burns, splenic disorders, and agents such as quinidine, sulfonamides, hydrochlorothiazide, diuretics, antiepileptics, and neuropharmacological agents have been implicated in the etiology of symptomatic thrombocytopenia. Decreases in platelet count also are accompanied by morphological changes in the size, shape, and cytoplasmic granulation of these cells and changes in adhesiveness and normal function in hemostasis and coagulation.

Studies on *platelet aggregation* have been of significant value in the study of platelet abnormalities and their role in disease states. The rate and extent of the aggregation and clotting response to adrenaline, ADP, collagen, and thrombin have been measured by observing changes in optical density of platelet-rich plasma on adding of these agents or other test substances. Low amounts of ADP give reversible aggregation, whereas a biphasic-aggregation pattern occurs with intermediate concentrations of ADP or with epinephrine. The second phase is the release of the platelets' endogenous ADP. High concentrations of ADP result in an irreversible aggregation. Aspirin acts as an inhibitor of the intrinsic-platelet ADP and the collagen reaction.

RETICULOCYTES

In normal peripheral blood 0.5% to 1.5% of the erythrocytes possess a fine reticulum in the cytoplasm. In blood smears prepared with Wright's, Giemsa, and other Romanowsky methods,

basophilic stippling of the erythrocytes occurs in lead poisoning (*plumbism*). This is not to be confused with the basophilic staining of the reticulocyte, which can only be seen when cells are stained by supravital procedures (mixture of dyes with wet blood prior to preparing of an air-dried blood smear). The observed granular filaments or reticulum of this immature erythrocyte are a result of endoplasmic coagulation by lipophilic dyes used in the supravital procedures. *Reticulocytes* can be enumerated manually by supravital staining of fresh blood with an anticoagulant-dye solution.

The usual method of expression is

$$\% \text{ Retics} = \frac{\text{No of reticulocytes/1000 RBC}}{10}$$

The "corrected" reticulocyte count is calculated for a more meaningful clinical approach in the degree of anemia by expressing the percentage of reticulocytes per microliter of whole blood.

$$\begin{matrix}\text{Corrected} \\ \text{reticulocyte} \\ \text{count}\end{matrix} = \begin{matrix}\text{Reticulocyte} \\ \text{count}\end{matrix} \times \frac{\text{(Patient's hematocrit)}}{\text{(Normal hematocrit)}}$$

In indirect counting methods a thin film of the blood-dye mixture is prepared on a microscope slide, counterstained with Wright's stain, and the reticulocytes enumerated in proportion to a predetermined erythrocyte count. In direct procedures, reticulocytes are enumerated in wet films without counterstaining. Suitable dyes are brilliant cresyl blue, methylene blue, and Janus green. These methods are subject to a high counting error.

Flow cytometric analysis using RNA-staining fluorescent dyes has improved greatly the precision and accuracy of reticulocyte counts. Dedicated analyzers such as the Sysmex R-3000 Reticulocyte Analyzer (*TOA Medical Instruments*) examine hundreds of thousands of cells during a 45-sec counting cycle and can subclassify reticulocytes by age.

An increase in the number of reticulocytes is an index of accelerated hematopoiesis and is observed in acute hemorrhage or adequate therapeutic management of iron-deficiency or pernicious anemia. In cases of chronic blood loss or bone-marrow depression a decrease in reticulocytes is seen.

BLOOD-VOLUME AND ERYTHROPOIETIC MECHANISMS

The mean red-cell mass in normal males is 2095 ± 384 mL (30 mL/kg), the average plasma volume is 2766 ± 459 mL (40 mL/kg), and the total blood volume is 4861 ± 795 mL (70 mL/kg). The specific determination of *red-cell mass* is estimated accurately by tagging erythrocytes with ^{51}Cr *in vitro* or ^{59}Fe *in vivo*. These isotopes are incorporated into the β-polypeptide (Cr) or porphyrin (Fe) of hemoglobin in the RBC and subsequent isotope dilution in blood after injection of tagged erythrocytes is used for calculation of red-cell mass. In hemolytic anemia, there is also a decrease in the normal life span (108–120 days) of the erythrocyte as indicated by a decreased survival time of ^{51}Cr-tagged red cells in blood.

Plasma volume is estimated by measurement of hemodilution of intravenous-injected ^{125}I or ^{131}I human serum albumin. The activity of labeled albumin steadily decreases after injection due to the loss of albumin to the extravascular space. Estimates of zero-time radioactivity levels can be made by extrapolation of a typical first-order blood-level decay curve. Dyes (Evans Blue) and other isotopes are less satisfactory for accurate assessment of plasma volume. The total blood volume is equal to the red-cell mass and plasma volume.

Chronic expansion of the red-cell mass is seen in primary and secondary polycythemia associated with erythrocytosis due to hypoxia, tumors, and renal disease. In these conditions, there is an increased hemoglobin and hematocrit and absolute increase in red-cell mass. In relative polycythemia, the high hematocrit is due to contraction of the plasma volume. *Chronic expansion of the blood volume,* with a resultant decrease in hematocrit value, and in some cases a "hemodilution" anemia,

is seen in cardiac failure, normal pregnancy, hepatic cirrhosis, splenomegaly, and arteriovenous fistula.

The metabolic defect in *pernicious anemia*, characterized by inadequate gastrointestinal absorption of vitamin B_{12}, is diagnosed readily by monitoring urinary radioactivity following oral administration of cyanocobalamin-[57]Co with and without intrinsic factor. The percent recovery of the isotope in normal patients is 3% to 25% and in pernicious anemia 0% to 2.5%.

Erythrocytes tagged with ^{51}Cr also are used in studying the effects of various compounds, such as the nonsteroidal anti-inflammatory drugs, on *gastrointestinal (GI) bleeding*. The patient's blood cells are tagged with ^{51}Cr and the agent under test is administered. If GI bleeding occurs, there is an increase in the ^{51}Cr content of fecal samples as a result of blood loss into the lumen of the GI tract.

Measurement of the absorption of radioactive iron (^{59}Fe), its tissue distribution (liver, spleen, precordium, sacral bone marrow), plasma elimination, and urinary excretion establish various *ferrokinetic parameters*. Iron is absorbed to the greatest extent as the ferrous salt in the upper small intestine. Absorption is decreased in iron overload, erythropoiesis, and various malignant, inflammatory, or infectious diseases. Iron is transported in plasma bound to transferrin, a specific iron-binding protein. Alterations in plasma iron and iron-binding capacity are seen in pregnancy, thalassemia major, and iron deficiency (hypochromic) anemia. Iron is stored in the liver, bone marrow, skeletal muscle, and spleen as ferritin and hemosiderin. The daily turnover of iron is about 35 mg, primarily from an "erythropoietic labile pool" in bone marrow.

Hemosiderosis is simply an increase in iron storage, whereas *hemochromatosis* denotes increased iron storage with associated tissue damage. Both of these states can result from oral or parenteral medicinal/transfusion iron overload. Iron excretion is limited and occurs by desquamation of iron-containing cells from the bowel, skin, and urinary tract. Iron-deficiency anemia is a symptom and not a disease. Treatment is based on evaluation of ferrokinetic parameters, correction of hemoglobin and tissue-iron deficiency, and recognition of the underlying cause (eg, chronic blood loss).

BLOOD COAGULATION

Hemostasis, the arrest of blood flow from a vessel, is regulated by extravascular (muscle, skin, and subcutaneous tissue), vascular (blood vessels), and intravascular (platelet-adhesion, clot-retraction, and blood-coagulation) mechanisms. The following discussion will be limited to those processes related to the blood-coagulation mechanism. When blood is allowed to clot, the free-flowing liquid is converted into a firm cell clot surrounded by serum. If an anticoagulant is added to blood, coagulation does not occur and the blood cells are suspended in a liquid phase—plasma. The clotting mechanism involves three stages: the formation of plasma *thromboplastin*, the conversion of *prothrombin* to *thrombin,* and the conversion of *fibrinogen* to *fibrin*.

The International Committee on Nomenclature of Blood Clotting Factors has numerically designated the blood-coagulation factors (Table 32-2). Fibrinogen and Factors V and VIII are absent in normal blood serum as a result of the clotting process. The absorption characteristics of certain blood-coagulation factors on calcium phosphate or barium sulfate are used in the differential analysis of specific factors. The interaction of coagulation factors may be initiated through either the intrinsic or extrinsic pathways. In the intrinsic system all the factors are present in the blood, while the extrinsic system is activated by the release of tissue thromboplastin. Figure 32-3 shows the activities of both pathways to form a stabilized fibrin clot.

In Stage 1 of the coagulation process, the contact of injured tissue with blood results in the activation of Factor XII, which reacts with calcium, plasma thromboplastin antecedent (PTA, Factor XI), plasma thromboplastin component (PTC, Factor IX), antihemophilic globulin (AHG, Factor VIII), and Factors III, V, and X to yield intrinsic or blood thromboplastin. This

Table 32-2. Blood-Coagulation Factors

| FACTOR	SYNONYM
I	Fibrinogen
II	Prothrombin
III	Thromboplastin (tissue)
IV	Calcium
V	Labile factor, proaccelerin, Ac globulin
VI	Accelerin
VII	Stable factor, proconvertin, serum prothrombin conversion accelerator (SPCA)
VIII	Antihemophilic globulin (AHG)
IX	Christmas factor, plasma thromboplastin component (PTC)
X	Stuart-Prower factor
XI	Plasma thromboplastin antecedent (PTA)
XII	Hageman factor
XIII	Fibrin-stabilizing factor (FSF)

stage normally is completed in 3 to 5 min. Extrinsic or tissue thromboplastin is formed rapidly (<12 sec) in various tissues in the body such as lung and brain in the presence of calcium and Factors V, VII, and X.

In Stage 2, thromboplastin catalyzes the conversion of prothrombin to thrombin (8–15 sec) in the presence of Factors V, VII, X, and calcium.

In Stage 3, the thrombin rapidly converts fibrinogen into fibrin, which then forms a network of fibers that traps red cells and thus forms the blood clot.

Although the exact nature of the enzymatic sequences in the coagulation process is not clear, it is definitely a biological amplification process starting from the small reaction of tissue contact to rapid conversion of fibrinogen to fibrin.

Blood contains natural inhibitors of coagulation such as antithrombin, heparin, and antithromboplastin, which can prevent a particular reaction in the coagulation sequence. The dissolution of blood clots occurs by the action of the blood proteolytic enzyme—plasmin or fibrinolysin. Plasmin is formed
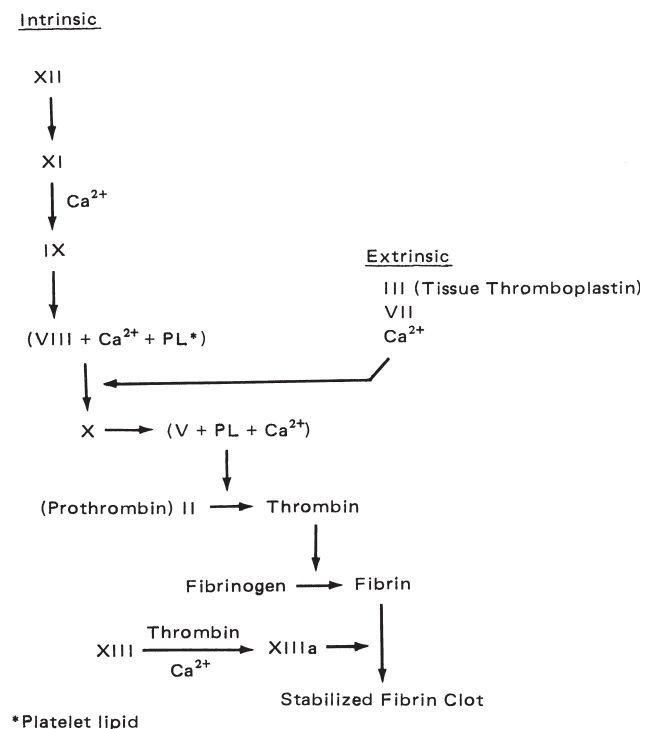
Figure 32-3. Blood coagulation process.

from its precursor, plasminogen, after activation by tissue and body fluids or substances of bacterial origin (streptokinase).

The routine tests performed in the coagulation laboratory are indices of vascular function (vascular phase and platelet adhesion) or intrinsic clotting mechanisms. Determinations of *bleeding time* and *capillary fragility* provide estimates of blood coagulation in the presence of platelets and tissue or vascular factors. In the Ivy method for determination of *capillary bleeding time,* a blood pressure cuff is placed on the forearm and inflated to 40 torr; a puncture wound is made and the time required for bleeding to stop is noted. *Bleeding time* is a screening test for disorders of platelet function or vascular defects but is usually normal in coagulation disorders. The test is useful in the differential diagnosis of Von Willebrand's disease (reduced factor VIII, with a normal bleeding time) from mild hemophilia. The normal bleeding time, as determined by this method, is 1 to 9 min. Dextran, pantothenyl alcohol and derivatives, penicillin G, nonsteroidal anti-inflammatory drugs, and streptokinase-streptodornase may cause a prolonged bleeding time. The Simplate 11 (*General Diagnostics*) is a standardized, disposable, spring-loaded bleeding-time device for platelet function testing. It uses two blades that are released automatically to produce two uniform incisions 6 mm long × 1 mm deep, making the procedure reliable and reproducible.

The *capillary fragility* or *tourniquet* test is based on the incidence of petechiae (small red marks) formation produced by an inflated blood pressure cuff over a 5-min period. Normally, a few tiny petechiae may appear. The most common cause of abnormalities in vascular-function and platelet-adhesion tests is thrombocytopenia.

An analysis of the *intrinsic coagulation mechanism* is concerned with the determination of the levels of the specific clotting factors in whole blood. In preliminary studies of a suspected hemorrhagic disorder, determinations of *coagulation time, clot retraction, platelet count, bleeding time,* and *capillary fragility* usually are performed.

In the Lee-White procedure, the coagulation time of whole blood is determined in regular or silicone tubes. Normal values are 8.5 to 15 min in glass and 19 to 60 min in silicone tubes. Anticoagulants and tetracyclines may cause increased times whereas corticosteroids and epinephrine cause decreased values. The siliconization of glassware prevents platelet aggregation, and thus delays coagulation. The samples used in the analysis of coagulation time are then inspected at 0.5, 1.0, 2.0, 4.0, and 24.0 hr after clotting to determine the time required for the various phases of clot retraction. The tubes also are observed for evidence of clot lysis or dissolution. The clot normally will start to retract in 30 min, completely retract within 24 hr and show no evidence of lysis over a 72-hr period. Prolonged coagulation times are associated with hemophilia, hypofibrinogenemia, and Factor IX deficiency. Abnormalities in any of these tests indicate the requirements for further coagulation studies.

The *prothrombin time test* is a measure of the levels of all coagulation factors, except III, IV, and VII, and is an index of the capacity of plasma to form thrombin. In the "One Stage" test, the plasma sample is mixed with calcium chloride and tissue thromboplastin, and the time required for fibrin-clot formation is determined. Results are compared with a normal plasma control, and the prothrombin time is reported either in seconds or as the percent of prothrombin calculated from a standard activity curve. Correction studies using normal serum, adsorbed normal plasma, or whole normal plasma added to test serum indicate deficiencies of Factors VII and X, Factor V, and Factor II, respectively. If none of these additives shorten the prothrombin time, a circulating anticoagulant problem can be suspected.

A modification of this technique (the *prothrombin-proconvertin procedure*) using a 1:10 dilution of both patient and control plasma in the presence of prothrombin-free plasma as a source of Factors I and V, is a more sensitive index of specific deficiencies in prothrombin, Factor VII, IX, and X.

Owren's *thrombotest*, as performed on whole blood, is sensitive to changes in both extravascular and intravascular clotting mechanisms, including Factor IX. The dosage of anticoagulant drugs, such as dicumarol, is adjusted in accordance with prothrombin-time determinations; patients are maintained usually within a therapeutic range of 20 to 40% of prothrombin activity (normal range, 80–130%). Reduced prothrombin levels, with prolonged prothrombin times, are observed in vitamin K deficiency, hemorrhagic disease of the newborn, excessive anticoagulant therapy, and liver and biliary disease. The interaction of other drugs with anticoagulants may cause increased prothrombin times. Drugs such as salicylates, phenylbutazone, oxyphenbutazone, indomethacin, and some sulfonamides increase the amount of active anticoagulant activity. Other drugs decrease the amount of vitamin K produced by gut bacteria, including chloramphenicol, kanamycin, neomycin, streptomycin, and the sulfonamides.

The *prothrombin consumption* test is an index of the efficiency of conversion of prothrombin to thrombin in the coagulation process. The blood sample is allowed to clot under standardized conditions and then the quantity of prothrombin complex removed in the serum is determined in the presence of extrinsic fibrinogen. At least 80% of the prothrombin is consumed normally. Reduced consumption of prothrombin (<80%) is observed in coagulation deficiencies (hemophilia) related to thromboplastin generation.

Other types of coagulation tests detect deficiencies in *thromboplastin generation mechanism*. The *thromboplastin generation time test* (TGT) provides a means of detecting specific deficiencies of Factors V, VIII, IX, X, XI, or XII. In the initial phase of this procedure, the clotting time of the patient's adsorbed plasma is determined in the presence of a standardized platelet factor reagent, calcium chloride, plasma substrate reagent (Factors I, II, and V), and the patient's serum. If the clotting time is abnormal (>16 sec), further tests are performed with the patient's plasma or serum. The adsorption of the plasma sample on barium sulfate removes Factors II, VII, IX, and X and facilitates differentiation of a Factor IX to X from V to VIII deficiency in the thromboplastin-generation mechanism. Thromboplastin generation is reduced in hemophilia and thrombocytopenia.

The *activated partial thromboplastin time* (PTT) *test* is based on the observation that hemophilic plasma has a normal clotting time in the presence of a complete thromboplastin (extrinsic-saline extract of brain tissue), as used in prothrombin determinations, but will give a markedly prolonged clotting time with an incomplete thromboplastin (cephalin). Cephalin is a thromboplastic, ether-soluble phospholipid factor with platelet-like activity. In this test, the clotting time of the patient's plasma is determined in the presence of calcium chloride and activated cephalin. This test is used primarily to detect deficiencies in Stage 1 of the coagulation mechanism and is rather sensitive to changes in Factors VIII and IX, as seen in classic hemophilia and Factor IX deficiency (Hemophilia B or Christmas disease).

In Stage 3 of the coagulation process, the presence of adequate levels of fibrinogen and thrombin is critical. *Fibrinogen levels* are analyzed semiquantitatively by determining the clotting time of a diluted plasma sample in the presence of extrinsic thromboplastin. This test is basically independent of prothrombin levels. Fibrinogen concentrations of 125 mg/dL or greater are adequate; deficiencies (hypofibrinogenemia) have been observed in liver disease, carcinomatosis, and in certain complications of pregnancy.

Increased levels of *fibrinogen degradation products* (FDP) have been demonstrated in serum due to primary activation of the fibrinolytic system (pathological fibrinolysis) or by secondary activation following increased blood clotting (disseminated intravascular coagulation). Fibrinogen (mol wt 3.4×10^5 daltons) is degraded sequentially to fragments X, Y, D, and E with molecular weights of 2.7, 1.65, 0.85, and 0.55×10^5 daltons, respectively. Fragments X and Y are more potent

anticoagulants than fragments D and E and are responsible for hemorrhagic states in defibrination. Complexes between fibrin monomer, fragment X, and other FDP interfere with thromboplastin generation and platelet formation. FDP can be measured by immunological techniques involving latex agglutination of particles sensitized with specific antibodies to FDP or by a hemaglutination-inhibition test. The normal level of serum FDP is 4.9 ± 2.8 μg/mL. Increased levels are seen in acute myocardial infarction, menstruation, complications of pregnancy, hypoxic newborns, malignancy, and renal disease.

Deficiencies in the clotting mechanisms usually can be corrected partially and temporarily by transfusion of normal blood or plasma. When this fails, the presence of *circulating anticoagulants* (antithrombin, antithromboplastins, heparin) must be considered. Heparin acts indirectly by means of antithrombin III, which neutralizes several activated clotting factors (XIIa, activated Fletcher factor, XIa, IXa, Xa, IIa, and XIIIa). The pharmacological effect of an oral anticoagulant is the inhibition of blood clotting by interfering with vitamin K–dependent clotting factors II, VII, IX, and X. Circulating anticoagulants are detected by determining the effect of normal plasma on the clotting time (*recalcification time*) of the patient's oxalated plasma in the presence of calcium chloride. If the addition of the normal plasma does not shorten the prolonged recalcification time, a circulating anticoagulant state can be reported.

Because the end point of all coagulation tests is the conversion of fibrinogen to fibrin, it is vital that analysts rigidly standardize their concepts of fibrin formation in visual recording procedures. The use of mechanical instrumentation in the detection of clot formation has increased significantly the standardization, accuracy, and reproducibility of coagulation procedures. These instruments measure and record the process of fibrin formation via increased turbidity (coagulogram or photometric clot detection) or changes in electrical conductance in the reaction mixtures. As well as performing routine coagulation tests simultaneously or sequentially, updated systems can run Fibrinogen and Factor assays, achieving rapid throughput and accuracy. New performance features are available with many of the automated coagulation instruments. These include precise temperature regulation, digital displays, automatic dilutions of patient samples, and the ability to measure specific clotting factors using chromogenic substrates.

Hemophilia is a classic deficiency of AHG (Factor VIII), Christmas disease of PTC (Factor IX), and Hageman trait of Factor XII. Hereditary or acquired deficiencies of Factors II, V, VII, X, and XI also are associated with disease states. The process of blood coagulation, analysis of coagulation factors, and interpretation of results comprise a highly complex system. The coagulation laboratory and the physician function together in the diagnosis and treatment of coagulation-deficiency diseases.

BLOOD-BANK TECHNOLOGY

Blood-bank technology in the modern laboratory is part of the blood-transfusion service. As whole blood for transfusion and its components are biologically active therapeutic substances, a complete analysis of their chemical and biological characteristics is vital to the assurance of successful therapeutic effects. The transfusion service is responsible for

1. Receiving and examining of the donor.
2. Collecting, processing, and storing the blood.
3. Typing of recipient and donor for ABO and Rh blood-group factors.
4. Compatibility (cross-matching) testing before transfusion.
5. Issuing of blood for transfusion and extracorporeal circulation.
6. Evaluating transfusion complications.
7. Performance of special serological tests pertinent to blood groups and other factors.

In this section a discussion of pertinent factors related to the various phases of the transfusion service will be presented.

RECEIVING AND EXAMINING OF THE DONOR

A complete registry[4] of prospective donors should be maintained, with specific reference to age, sex, weight, address, occupation, and telephone number. Computerized blood banking has increased the efficiency of this service. Donors should preferably be between the ages of 21 and 60 and should weigh no less than 110 pounds. The donor may be rejected on the basis of previous or active incidence of certain microbial diseases (recurrent malaria, syphilis, infectious or homologous serum hepatitis, tuberculosis), bleeding abnormalities, convulsions, allergic syndromes, skin or heart diseases, diabetes, alcohol or drug addiction, pregnancy, cancer, recent immunization with live vaccine product, acquired immune deficiency syndrome (AIDS), or blood pressure abnormalities (acceptable blood pressure: between 100/50 and 200/100; pulse rate: 60 to 120/min). The screening of blood for exposure to human immunodeficiency virus (HIV) is crucial to reducing the risk of infection from transfusion. ELISA (enzyme-linked immunosorbent assay) screening tests for the detection of antibodies against HIV are available from manufacturers. More sensitive tests are available to detect viral DNA in body fluids.

A period of at least 8 weeks should have elapsed since blood was withdrawn and the blood hemoglobin level should be 12.5 to 13.5 g/dL or greater. Serum bilirubin and transaminase levels also should be evaluated in donors with previous incidence of jaundice.

COLLECTING, PROCESSING, AND STORING THE BLOOD

A tourniquet is applied to the arm of the donor to occlude the venous return, the skin area is sterilized, and the blood is collected by venipuncture (phlebotomy). NIH Formula A or B ACD (Acid-Citrate-Dextrose) or ACD-phosphate solutions are used as anticoagulants in the sterile blood-collecting containers. Evacuated containers may be of regular or silicone glass; collapsible plastic containers offer many advantages in donation, blood-banking, and transfusion procedures.

The preservation of the red cells in blood is improved by the complete removal of trapped air in the blood-collection apparatus, rapid cooling after collection, and storage at 4°. Properly collected whole blood is usually stable for 21 days at 1° to 6°. The deterioration of whole blood is related to increased cellular fragility (increased plasma K^+) and decreased glucose utilization. Blood that is used for correction of any bleeding tendency or clotting defect should be as fresh as possible. Leukocytes, platelets, and Factors V and VIII deteriorate in stored plasma or whole blood.

ABO BLOOD-GROUP CLASSIFICATION[5]

Human red cells can be classified into various groups or types on the basis of reactivity of certain blood factors (*agglutinogens*) located on the erythrocyte membrane. The Landsteiner system (Table 32-3) for the four blood groups is based on the presence or absence of either A or B agglutinogen on the cell surface (Group A, B, AB, or O, respectively).

Serum does not contain the antibody (*agglutinin*-IgM type) for the antigen present in an individual's own red cells, but does contain the isoagglutinin (eg, anti-B in blood group A) due to exposure, early in life, to bacterial and plant antigens similar in structure to the A-B antigens. The clumping or agglutination of the red cells by reaction of agglutinogen with agglutinin is used in blood-grouping techniques. In certain instances hemolysin antibodies, present in serum containing anti-A or anti-B agglutinins, cause the disruption of cells and release of hemoglobin (hemolysis).

Human blood cells are grouped by two separate reactions: cellular or "front" grouping and serum or *reverse* grouping. The blood group ordinarily is determined by testing an individual's red cells with standardized anti-A or anti-B serum (certified by the Bureau of Biologics, FDA). Confirmation of the blood group

Table 32-3. Blood-Group Systems

BLOOD GROUP	AGGLUTINOGEN IN CELL	AGGLUTININ IN SERUM	REACTION[a] WITH ANTI-A SERUM	REACTION[a] WITH ANTI-B SERUM	FREQUENCY (%) IN CAUCASIANS
A	A	Anti-B	+	−	41
B	B	Anti-A-A_1	−	+	10
AB	AB	None	+	+	4
O	None	Anti-A and B	−	−	45

[a] Agglutination.

(reverse typing) is accomplished by an analysis of an individual's agglutinin titer. In this procedure the individual's serum is heated at 56° for 10 min to destroy hemolysins, and then mixed with known Subgroup A_1 or B_1 human red (Rh-negative) cells in the agglutination test. These two tests should be in agreement prior to the release of blood for transfusion.

Although human blood cells of Group B react uniformly with Anti-B serum, Group A and AB cells show a wide range of reactivity with Anti-A or Anti-A_1B serum. Blood-group A may be further categorized into Subgroups A_1, $A_{int.}$, A_2, A_3, A_0, and A_x on the basis of the reaction with absorbed Anti-A, Anti-A_1-lectin, Anti-H-lectin, Anti-$A_{1,2}$, and Anti-AB serum and the presence of Anti-A_1 in the serum. Certain Group O individuals possess anti-H in their serum and are further subcategorized into the Bombay or O_h phenotype. Tests for A, B, and H in saliva can establish the genotype of an individual, that is, A and H in saliva of blood-group A; B and H in B; H and O and A, B, H in AB. This is helpful in cases of poorly developed red-cell antigens or in the loss of cellular antigen in some patients with leukemia.

As the human blood cell contains many antigens with rather complex biochemical and immunochemical properties, the blood factors have been classified further into various subsystems. The Kell (K), Lutheran (Lu), Lewis (Le), Duffy (Fy), Kidd (Jk), MNS, Sutter (Js), Diego (Di), and P blood-factor systems are based on the detection of a specific antigen on or within the red cell by means of antibody (*isohemagglutinin*) reactions with specific antisera or panels of reagent red cells. Some of these factors (eg, Kidd, Kell, and Lewis) have been involved in transfusion reactions.

THE RH-HR SYSTEM AND ANTIHUMAN GLOBULIN TEST

The presence or absence of Rh_0 *antigen* in human blood is of prime importance in transfusion reactions, paternity disputes, and isosensitization phenomena. There are eight blood Rh phenotypes that are determined by their reaction with three specific serum agglutinins (Anti-Rh_0, Anti-rh′, and Anti-rh″): rh, rh′, rh″, rh′rh″, Rh_0, Rh'_0, Rh0″, and $Rh'_0Rh''_0$. The rh groups do not contain the Rh_0 factor on the cell surface and are designated "Rh-negative." The terminology of the Wiener system (Rh, rh) is comparable to the Fisher-Race (CDE) as follows: rh′(C), Rh_0(D), rh″(E). The Rosenfeld system uses a numerical classification: $Rh_1 = Rh_0$.

The absence of the Rh antigen in about 15% of the population does not preclude the presence of other factors; the use of specific antisera (Anti-hr′ and Anti-hr″) has demonstrated the existence of the Hr factors (Hr_0, hr′, hr″). For example, the Rh-negative cell (rh″) possesses rh″hr′Hr_0 antigens. The antigen Rh_0(D) is the most potent immunogen of all the Rh antigens.

The Rh antibodies are either *saline agglutinins* (complete) or "blocking" antibodies (incomplete). The latter are of the IgG type. They are used in Rh testing procedures and are produced more commonly, and in higher titer, in the human isosensitization or autoantibody reactions. They will not agglutinate saline suspensions of normal Rh-positive red cells except in the pres-

ence of a high concentration of albumin, serum or conglutinin (AB serum with albumin) at a temperature of 35° to 37°.

In routine Rh testing procedures, a sample of blood (oxalated or heparinized) or a suspension of cells in serum or albumin is mixed with Anti-Rh_0 serum on a slide or in a tube at 37° to 47°. The presence of clumping indicates that the blood possesses Rh_0 antigen. Confirmation of an Rh-negative test may be performed by retesting with Anti-rh′Rh_0rh″ serum.

In Rh testing procedures, red cells from patients with acquired hemolytic anemia are partially coated with human autoantibody, and cells from erythroblastic infants are coated with maternal antibody globulins and may be clumped falsely by Rh typing serum containing a high protein concentration, or may appear to be Rh-positive in the saline-cell suspension test. Demonstration of anti-Rh_0(D) in an eluate from these antibody-coated cells can help to establish true Rh type.

Anti-Rh antibodies are not normally present in human serum; they may be acquired via isosensitization. The transfusion of Rh-positive blood to an Rh-negative recipient, or transfer of cells of Rh-positive fetus through the placental barrier to the Rh-negative mother, will result in formation of antibodies to Rh agglutinogens not present in the cells of the recipient or mother, respectively.

Hemolytic blood-transfusion reactions and hemolytic disease of the newborn (erythroblastosis fetalis) involve *isosensitization phenomena* usually related to the Rh_0 antigen. Hr and ABO antigens also can be responsible for hemolytic disease of the newborn. If an expectant mother is Rh-negative and the father is Rh-positive, the Rh genotype of the father should be determined. If the father is homozygous, the erythrocytes will contain a pair of Rh_0 factors and the offspring will inherit the Rh_0 factor; if he is heterozygous, one Rh_0 and one Hr_0 factor will be present and his offspring may or may not inherit the factor.

If the fetus is Rh-positive, the mother may be sensitized to the Rh antigen and in subsequent pregnancies the development of high titers of Anti-Rh_0 antibodies will result in hemolytic disease of the fetus. These antibodies enter the fetal circulation via the placental barrier, coat the red cells of the fetus, and cause excessive erythrocyte destruction, hyperbilirubinemia, and associated potential for brain damage, hydrops fetalis (edema), and congenital anemia of the newborn. This Rh disease can be avoided now by proper therapeutic use of Rh_0(D) Human Immune Globulin (Rh_0-GAM, *Ortho*) to prevent the postpartum formation of active antibodies in the Rh_0 (D)-negative, D^u-negative mother who has delivered an Rh_0(D)-positive or D^u-positive infant.

The *Coombs' antiglobulin test* is a method of detecting the blocking-type antibodies, globulins, and complement that are attached to red-cell antigens in isosensitization phenomena.
In the *direct* test procedure, a saline suspension of washed red cells is mixed with antihuman gamma globulin antiserum and agglutination is indicative of the combination of human antibody with antigen on the red cell, such as maternal incomplete isoantibody on infant's red cells in hemolytic disease of the newborn, autoimmune, drug-induced, alloantibody-induced hemolytic anemia, and after transfusion of incompatible red cells.

An *indirect* procedure is used to demonstrate the presence of blocking antibody in the serum of pregnant Rh-negative women and in transfusion reactions. In this procedure the patient's serum is incubated with a suspension of Group O Rh-positive red cells; the cells are washed and then antihuman globulin antiserum is added to detect the coating of the red cells with antibody globulin from the patient's serum by agglutination phenomena. If agglutination occurs in the first part of the procedure, a saline agglutinin is also present.

Anticomplement sera (anti-nongammaglobulin antiserum) are used to detect reactions involving anti-JK.

The Du allele is a clinically important variant of the Rh_0 factor and usually associated with rh′(C) and rh″(E). Individuals with this factor are considered Rh-positive; the red cells fail to react with anti-Rh_0 in the saline-tube method but react with incomplete anti-Rh_0(D) by other slide or tube techniques. Rh-

negative donors should be tested for Du factor. If positive, their blood must only be given to Rh-positive recipients.

DRUG-RELATED PROBLEMS

Hematological abnormalities may be caused by the administration of drugs that can cause a positive direct antiglobulin test and immune hemolytic anemia, such as cephaloridine, cephalothin (*Keflin*), methyldopa (*Aldomet*), penicillin, L-dopa, quinidine, phenacetin, and insulin.

COMPATIBILITY TESTING

Cross-matching procedures are designed to detect incompatibilities in the blood of donors and recipient. The test is designed to prevent transfusion reaction and assure maximum benefit to the patient. Although erroneous ABO grouping usually will result in an incompatible cross match, no such protection exists in the Rh system. An incorrectly typed Rh-positive donor blood can result in primary immunization to $Rh_0(D)$ antigen if transfused to an Rh-negative recipient. For each transfusion, a *major* and *minor cross match* should be performed.

In the *major cross match* (1) a saline suspension of the donor's cells is mixed with the recipient's serum and (2) the donor's cells are suspended in recipient's serum or in serum with added albumin. The saline cross match is an additional check on the ABO typing and may detect incompatibilities caused by antibodies to M, N, S, P, and Lu subgroups. The high-protein or albumin cross-match can demonstrate antibodies in the Rh system. The presence of agglutination or hemolysis indicates incompatibility.

The *minor cross match* includes the donor's serum and the recipient's cells and is useful as a check of the ABO typing and an indication of the possibility of transfusion reactions caused by a rare antigen on the recipient's cells or uncommon antibodies directed against an antigen in the serum of the donor. The minor cross match has been replaced in many instances with screening of the donor's serum against a panel or pool of red cells of known antigenicity.

The *indirect antihuman globulin* procedure also must be performed with the recipient's serum and donor's cells with and without albumin (major side) and may be tested with the donor's serum and recipient's cells (minor side). The use of proteolytic enzymes (bromelain) enhances the agglutination of red cells by low-titer or weakly reacting Rh-Hr antibodies, probably by removing sialic acid residues on the RBC surface. The red cells used in the indirect Coombs' test are treated with the enzyme prior to absorption of antibodies and addition of antiglobulin reagent.

The usual cross-matching techniques involve (1) a room-temperature or 30° procedure, preferably with the addition of albumin, (2) a high-protein procedure, and (3) an antiglobulin procedure.

The presence of nonspecific *autoantibodies, cold agglutinins,* and *bacteriogenic agglutination* sometimes complicates the cross-matching procedure. If the recipient's serum reacts more strongly with his or her own cells than with the donor's, autoantibodies should be suspected. Cold agglutinins usually will agglutinate all blood, regardless of type, at low temperatures, but will not react at 37°. Agglutination as a result of bacterial contamination of blood is called panagglutination.

HEPATITIS TESTING

Post-transfusion hepatitis is associated with the transmission of virus-like particles referred to as *Australia* or *serum hepatitis antigen* or the *hepatitis-associated antigen* (HAA). All donor blood must be tested for the presence of HAA. Agar gel diffusion (AGD), counterelectrophoresis (CEP), complement fixation (CF), and rheophoresis procedures can be used.[6] The rheophore-sis procedure uses a modified gel-diffusion technique for the detection of HAA by precipitin-type reaction with HAA antibody. It offers the sensitivity of CEP and CF procedures with the simplicity of the AGD procedure. Other tests for HAA are based on radioimmunoassay (RIA) technique for detection of antigen by hemagglutination (HA) or HA-inhibition for the presence of HAA antibody. In the RIA technique, the donor's serum is added to a test tube coated with HAA antibody (solid RIA). If the serum contains HAA, it will bind to the antibody. ^{125}I-HAA is then added to the tube. If the antibody binding site is occupied previously with HAA from the donor's serum, ^{125}I-HAA will not bind and the determination of ^{125}I bound versus free is an index of HAA content of the donor's serum.

ISSUING OF BLOOD AND EVALUATING TRANSFUSION REACTIONS

Whole-blood, red-cell, or leukocyte suspensions, plasma, platelet-rich plasma, platelet concentrates, leukocyte-poor blood, AHF, factor IX complex, plasma protein fractions, and RhoGAM are products of the transfusion service.[7] Transfusion reactions are related to antibody phenomena or disease transmission. The hemolytic reaction resulting from the transfusion of incompatible cells is the most serious problem. The transfusion of microbially contaminated blood can result in a pyrogenic reaction or transmission of infectious diseases, such as malaria, syphilis, AIDS, or hepatitis. Allergic reactions (urticaria, asthmatic seizures), circulatory overload, embolic complications (blood clot, air emboli) also may be encountered. Leukocyte and platelet antibodies develop in repeat transfusions and in transplantation patients. The transfusion service is an integral unit in evaluating such complications.

TECHNIQUES OF ANALYSIS

This section will describe the principles of the procedures used in the analyses of various substances in blood, plasma, or urine. Examples of the significance of such tests in clinical diagnosis will be presented. For a complete description of the physiological and pharmacological aspects of these blood constituents, see the *Bibliography*.

INSTRUMENTATION

The development of instrumentation has accelerated progress in clinical chemistry. An excellent review of the principles and applications in clinical chemistry of automation, atomic-absorption spectroscopy, ultraviolet and visible spectrophotometry, fluorimetry, phosphorimetry, infrared and Raman spectroscopy, microwave, and radiowave spectroscopy and nucleonics was prepared by Broughton and Dawson.[8] Quality-control techniques are a vital part of any clinical laboratory. Standard reference materials,[9,10] standardization of quantities and units,[11] and continual evaluation of precision and accuracy of various determinations[12] are incorporated into procedures of all reliable clinical laboratories. The manufacture of certified standards and reagents and the certification of clinical chemists and clinical laboratories are under the supervision of either the Food and Drug Administration (FDA), National Institutes of Heath (NIH), Pharmaceutical Manufacturers Association (PMA), American Association for Clinical Chemistry, the College of American Pathologists, and the National Committee for Clinical Laboratory Standards (NCCLS).

INTERACTION OF DRUGS WITH CLINICAL LABORATORY TESTS

Drugs may interfere with the interpretation of laboratory tests by three classes of mechanisms:

1. *Chemical or biochemical* interference due to reaction of a drug or its metabolite in biological fluids with test reagents in analytical procedures. Examples of Class 1 interference include false-positive urine glucose results due to the reducing properties of drugs or metabolites such as ascorbic acid, *p*-aminosalicylic acid, tetracycline, cephaloridine, and levodopa, which are excreted in urine. Spironolactone will result in an elevation of certain urinary ketosteroids through cross-reaction of the drug in the analytical procedure.

2. *Pharmacological* interference due to normal drug-induced alterations in various physiological parameters. Examples of Class 2 interference include the decrease in serum-potassium levels in patients receiving thiazide diuretics, the alteration in serum uric acid with probenecid, and the elevation in various plasma proteins and thyroid function tests with estrogen-progesterone combinations. Drug–drug interaction also can result in changes in these parameters. Guanethidine enhances the effect of the coumarin anticoagulants. Barbiturates induce hepatic microsomal enzyme synthesis and subsequently increase the metabolism and decrease the therapeutic effect of drugs, such as warfarin, even after these drugs are terminated.

3. *Toxicological* interference as a consequence of the toxicity of a drug. Examples of Class 3 interference include changes in liver- and kidney-function tests and hematological parameters (anemia, agranulocytosis, leukopenia) due to drug-induced toxicity and positive LE and ANA tests due to a "lupus-like" syndrome induced by hydralazine.

It is beyond the scope of this chapter to include a complete listing of drug interactions in laboratory tests. The reader is referred to an annual, readily available, computerized review of the effect of normal therapeutic drug doses, as well as overdoses, on clinical laboratory tests[13] and to other review articles.[14]

Blood

COLLECTION AND PREPARATION FOR CHEMICAL ANALYSIS

Using aseptic technique, a blood sample is obtained by venipuncture and usually drawn directly into evacuated glass tubes. The choice of anticoagulant, type of specimen, stability of test component, and use of preservatives depends on the type of analysis requested and the specific analytical procedure involved. If serum is desired, the blood sample is allowed to clot and the serum is separated by centrifugation. When whole blood or plasma is to be used in the analysis, an anticoagulant is added to the collecting tube.

The following concentrations of specific anticoagulants are used routinely per 10 mL blood: lithium, potassium, or sodium oxalate (15–25 mg), sodium citrate (40–60 mg), heparin sodium (2 mg), disodium or tripotassium ethylene-diaminetetraacetate (EDTA-Na$_2$, 10–30 mg), or ACD-Formula B solution (1.0 mL).

Heparin prevents blood coagulation by inhibiting the thrombin-catalyzed conversion of fibrinogen to fibrin. The other anticoagulants either precipitate blood calcium or convert ionized calcium into a nonionized (chelated) form that cannot function in the coagulation reaction. Heparin and EDTA do not alter the cellular elements of blood significantly. Sodium fluoride and thymol are used as preservatives or enzyme inhibitors to prevent the deterioration of various substances in the blood sample; for example,

<p align="center">glucose → lactic acid.</p>
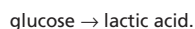
Preservatives and anticoagulants can interfere with some enzyme tests. Serum usually is used for these procedures.

The separation of plasma or serum, and chemical analysis, usually are performed as soon as possible after the collection of the sample. The addition of polystyrene granules to the blood sample prior to centrifugation facilitates the isolation of serum or plasma. Hemolysis interferes with analytical procedures for bilirubin, albumin, nonprotein nitrogens, pH, phosphorus, potassium, and various enzymes. The serum also should be observed for presence of lipemia. Changes in the ratio of CO$_2$, chloride, and electrolytes in cells and plasma, glycolytic conversion of glucose to lactic acid, hydrolysis of ester phosphate to free inorganic phosphate, bacterial conversion of urea to ammonia, and conversion of pyruvate to lactate are examples of changes that can occur in contaminated, improperly preserved, or unrefrigerated blood specimens.

The first stage in many of the classic manual chemical determinations is the removal of blood protein and preparation of *protein-free blood filtrate*. The protein is precipitated with tungstic acid, trichloroacetic acid, zinc hydroxide, or organic solvents, such as alcohol and acetone, and then filtered or centrifuged to remove the protein coagulum. Tungstic acid precipitation is performed by mixing 1 volume of blood or 2 volumes of plasma with 9 volumes of stabilized tungstic acid reagent. The filtrate obtained in this procedure should be in the pH range of 3.0 to 5.1 to assure the adequate removal of proteins (<2 mg/dL in filtrate).

The Somogyi filtrate is prepared by mixing 1 volume of blood with 5 volumes of water, 2 volumes of 5% zinc sulfate, and 2 volumes of 0.3 *N* barium hydroxide. The barium sulfate is precipitated and the zinc hydroxide formed in the reaction precipitates the blood proteins. Trichloroacetic acid (10%), in a ratio of 9:1 with blood, yields greater volumes of filtrate due to a more complete formation of protein agglomerates.

BLOOD GLUCOSE

Methods for determining blood glucose are based on the use of glucose as a reducing agent or on the enzymatic oxidation of glucose to gluconic acid. In the Folin-Wu technique, glucose is determined in a protein-free blood filtrate by reduction of alkaline cupric sulfate and subsequent reaction with phosphomolybdic or arsenomolybdic acid reagent to form a blue complex that can be estimated colorimetrically. The Nelson-Somogyi method uses a protein-free blood filtrate prepared with zinc hydroxide to remove most of the interfering reducing substances.

The presence of a terminal aldehyde in the glucose molecule is the basis of a colorimetric determination with phenolic hydroxyl reagents (phenol in aqueous methyl salicylate or phosphorylated 1,3-dihydroxybenzene) in the presence of strong sulfuric acid and heat. The *o*-toluidine procedure is a color reaction specific for hexoses—glucose, mannose, and galactose. Because aldohexoses other than glucose are normally present in very small concentrations, results obtained by this method approach the true value of glucose. *o*-Toluidine is condensed with glucose in glacial acetic acid to yield a green chromogen by forming an equilibrium mixture of a glycosylamine and Schiff base.

In the preceding techniques, interfering substances such as lactose, galactose, and glutathione are measured, and the value is reported in the nonspecific term "sugar." Enzymatic determination with glucose oxidase is the only test specific for blood glucose. Blood glucose is converted to gluconic acid and hydrogen peroxide by glucose oxidase; the peroxide is then estimated by iodimetric procedures or by oxidation of a chromogen (*o*-dianisidine or 2,2′-azino[diethylbenzothiazolinesulfonic acid]) in the presence of a peroxidase to form a colored product. Drugs that cause a slight increase in glucose values include ACTH, corticosteroids, D-thyroxine, diazoxide, epinephrine, estrogens, indomethacin, oral contraceptives, lithium carbonate, phenothiazines, phenytoin, thiabendazole, and diuretics. Drug interferences with *o*-toluidine methods, which cause a slight increase, include ascorbic acid, dextran, fructose, galactose, mannose, ribose, xylose, and bilirubin.

Another enzymatic procedure uses the hexokinase-catalyzed conversion of glucose to glucose 6-phosphate (G6P), and then to 6-phosphogluconate and nicotinamide-adenine-dinucleotide phosphate (NADPH) in the presence of NADP and G6P dehydrogenase. The NADPH thus formed is equivalent to the amount of glucose present and is estimated spectrometrically at 340 or 366 nm.

Normal fasting blood-sugar values for adults are 80 to 120 mg/dL; true glucose is 65 to 100 mg/dL. When the blood-sugar

values exceed 120 (hyperglycemia), diabetes mellitus should be suspected and can be confirmed by evidence of diminished carbohydrate tolerance. The effect of ingested carbohydrate on blood sugar can be determined by the *glucose tolerance test;* 100 g of glucose (1.75 g/kg) in water or a flavored beverage is administered orally, and glucose determinations are performed on blood and urine samples at hourly intervals for 3 hr. Values above 160 at 1 hr and 110 at 2 hr in blood samples are abnormal. The renal threshold for glucose is 180 to 200 mg/dL of blood, and thus sugar should not appear in the urine of normal subjects in the tolerance test.

Hyperglycemia and decreased glucose tolerance are seen in diabetes mellitus (to 500 mg/dL) and hyperactivity of the adrenal, pituitary, and thyroid glands. *Hypoglycemia,* with a blood-sugar value of <60 mg/dL and increased glucose tolerance, is encountered in insulin overdose, glucagon deficiencies, and hypoactivity of various endocrine glands. Intravenous glucose tolerance studies are used to circumvent defective absorption of glucose in the GI tract, for example, in steatorrhea.

Monitoring hemoglobin A_{1c} is another way to follow patients with hyperglycemia. This is more specific for diagnosing diabetes but less sensitive than the glucose tolerance test.[15] Normally, hemoglobin A_{1c} accounts for 3% to 6% of the total hemoglobin, whereas in diabetics it is 6% to 12%. The concentration of Hgb A_{1c} in the blood reflects the patient's carbohydrate status over a period of time, providing a marker for hyperglycemia. *Pancreatic function tests* include studies on intravenous and oral glucose, glucagon, and tolbutamide tolerance. The beta cells of pancreatic islet tissue secrete insulin and the alpha cells secrete glucagon, a substance antagonistic to insulin and having a hyperglycemic effect induced by its glycogenolytic action. In *glucagon tolerance studies,* the effect of parenteral administration of glucagon on blood-sugar values is useful in the diagnosis of pancreatic and hepatic function. *Insulin and tolbutamide tolerance studies* are used in the diagnosis of endocrine disorders, differentiation of insulin-resistant diabetics, and determination of functional hypoglycemia and islet-cell tumors.

Galactosemia, the presence of galactose (>4.5 mg/dL) in blood, is usually due to an inborn error of galactose metabolism. Congenital deficiencies in galactokinase or galactose 1-phosphate uridyl transferase result in inadequate galactose metabolism with accumulation of galactose 1-phosphate in the liver. Oral administration of galactose in galactosemia leads to a decrease in blood glucose and an increase in concentrations of galactose in the urine and blood. Galactose is measured by estimation of NADH liberated in the conversion of galactose to galactonolactone in the presence of nicotinamide-adenine dinucleotide (NAD) and galactose dehydrogenase. Deficiencies in intestinal disaccharidases such as lactase will preclude efficient conversion of lactose to galactose and glucose, and oral administration of lactose will cause no increase in blood galactose and usually produce diarrhea. Galactose-loading studies are useful in the diagnosis of toxic or inflammatory conditions of the liver. In hepatic cirrhosis, there is a decrease in the galactose-metabolizing capacity of the liver due to the inhibition of hepatic diphosphogalactose-4-epimerase.

Lactic acid is a product of glucose metabolism; it is converted into pyruvic acid and NADH by lactate dehydrogenase (LDH) in the presence of NAD. Blood lactic acid is estimated by reaction with LDH to form pyruvate and NADH; the NADH level is determined spectrophotometrically at 340 nm and is a function of lactic acid concentration. It is elevated (>20 mg/dL) following exercise, anesthesia, and certain types of acidosis. The *blood lactate/pyruvate* ratio should be calculated to determine the presence of excess lactic acid in the blood in acidosis, thiamine deficiency, and decompensated heart disease.

Blood pyruvic acid is determined by the reverse procedure, that is, the conversion of pyruvate to lactate in the presence of LDH and NADH. Normal blood pyruvic acid ranges from 0.6 to 1.3 mg/dL by chemical methods and 0.3 to 0.7 mg/dL by enzymic procedures.

NONPROTEIN NITROGEN COMPOUNDS

Nonprotein nitrogen (NPN) compounds refer to all nitrogen-containing compounds in biological fluids exclusive of protein, including nitrogen from amino acids, low-molecular-weight peptides, urea, nucleotides, uric acid, creatinine, creatine, and ammonia. Blood NPN usually is determined by digesting a protein-free blood filtrate with sulfuric acid in the presence of a catalyst (SeO_2) to convert nitrogen to ammonium sulfate (Kjeldahl digestion); the excess acid is neutralized and ammonia determined by Nesslerization or reaction with alkaline hypochlorite.

The normal blood NPN is 25 to 45 mg/dL (48% urea N, 14% amino acid N, 4% creatine N, 1% creatinine N, 3% uric acid N, and 30% residual N). In renal damage, NPN is elevated to values ranging from 60 to 500 mg/dL (*azotemia*). As variations in NPN mainly reflect alterations in blood urea nitrogen (BUN), urea determinations are more sensitive and preferred as a guide to kidney function.

The primary pathway of nitrogen metabolism in man is the synthesis of urea from ammonia in the liver and then rapid renal excretion of urea. In renal disease (*nephritis*), the excretion of urea is diminished, and blood NPN and BUN are increased. In BUN procedures, *urea* is converted enzymatically to ammonia by urease; the ammonia then is determined by Nesslerization, reaction with phenol-alkaline hypochlorite, aeration into standard acid and subsequent titration or reaction with salicylate-nitroprusside reagent at pH 12 in the presence of alkaline dichloroisocyanurate to form a green chromogen that can be estimated colorimetrically. The ammonia also can be estimated by spectrophotometric determination of NAD produced in the conversion of ammonia and α-ketoglutarate to glutamate by NADH-L-glutamate dehydrogenase. Direct chemical determinations of urea are based on the reaction with 2,3-butanedione in an acid medium (Fearon reaction).

BUN (normal = 5–25 mg/dL) is increased in chronic and acute nephritis, metallic poisoning, and cardiac failure; reduced levels occur in rapid dehydration or following diuresis. In severe liver damage due to diminished urea formation, an increase in blood ammonia and decrease in BUN are observed. Urine urea output (6–17 g/day) is an index of *glomerular filtration rate* (GFR) *and kidney function*. Increased dietary protein and gastrointestinal hemorrhage will increase urine urea. Decreases in urea excretion involve either tubular reabsorption or secretion defects.

The *nitrogen balance* represents the balance between nitrogen input or produced (N_{in}) and nitrogen excreted (N_{out}); in normal individuals $N_{in} = N_{out}$. N_{out} is regulated by renal GFR; in renal disease GFR is decreased, $N_{in} > N_{out}$, and BUN is increased. The rate of urinary excretion of parenterally administered dyes (phenolsulfonphthalein), inulin sodium, *p*- aminohippurate, and mannitol are sensitive indices of GFR in *renal clearance studies*.

Creatine (methylguanidoacetic acid) and *creatinine* (creatine anhydride) are involved in the physiology of muscle contraction. Creatine phosphate is an intracellular source of high-energy phosphate bonds via the reaction of adenosine triphosphate (ATP) and creatine kinase. Creatinine is the waste product of creatine metabolism and is the normally excreted compound.

Serum creatinine is determined by reaction with alkaline picrate to form a red chromogen. These values usually represent 20% to 30% of noncreatinine-interfering substances. Absolute determinations can be made by the absorption of creatinine from protein-free blood filtrates on aluminum silicate prior to the final determination. Drugs causing nephrotoxicity result in a slight increase in creatinine, and those that interfere with color formation in the reaction include bromsulfophthalein (BSP), phenolsulfonphthalein (PSP), acetoacetate, ascorbic acid, levodopa, methyldopa, glucose, and fructose. Creatine is determined after hydrolytic conversion to creatinine with boiling, aqueous picric, or hydrochloric acid.

Renal clearance of endogenous creatinine is related to GFR and is normally 1 to 2 g/day (creatinine coefficient = 20–26 mg/kg/24 hours). Normal serum creatinine is 1 to 2 mg/dL; creatine 0.2 to 1.0 mg/dL. Higher values (5 mg/dL) indicate glomerular damage or cardiac insufficiency.

Uric acid is a catabolite of purine metabolism as derived from nucleic acids or nucleotide cofactors. Direct methods for determining uric acid involve the reaction with alkaline phosphotungstic acid to form a "tungsten blue," which is estimated colorimetrically. In another method, alcoholic NaOH is added to a protein-free filtrate to eliminate interfering reducing substances (ascorbic acid, glutathione) prior to the reduction of uric acid with acid copper chelate to form a cupric chromogen complex.

In indirect procedures, uric acid is hydrolyzed by the enzyme uricase; the decrease in absorbance at 290 to 293 nm is a function of the initial concentrations of uric acid. The normal blood value is 1.5 to 6.0 mg/dL. It is elevated in renal disease, gout due to increased metabolic pools of uric acid, and leukemia as a result of increased turnover of cellular nucleoprotein.

Amino acid determinations in blood are performed by conventional colorimetric ninhydrin techniques or reaction with alkaline β-naphthoquinone-4-sulfonate. Normal plasma values range from 3.9 to 7.8 mg/dL. A variety of metabolic disorders may be detected by analyzing for increased levels of specific amino acids in the urine or blood. Total urine amino acids are determined by formol titration; formaldehyde reacts with basic amino groups and thus permits subsequent titration of the acidic groups of the amino acids. Daily excretion of amino acid nitrogen ranges from 100 to 400 mg, constituting 1% to 2% of total urine nitrogen.

The identification and quantitation of specific amino acids in the blood and urine are accomplished by paper, thin-layer (TLC), column, and ion-exchange chromatographic and electrophoretic separation of electrolytically desalted blood or urine samples (see Chapter 33).

Abnormal amino acid metabolism (*aminoacidopathies*) usually results in the presence of abnormal quantities of specific amino acids in the urine (aminoaciduria). The aminoacidurias are divided into two main groups:

1. *Primary overflow aminoaciduria* in which blood amino acids are elevated phenylketonuria (PKU), maple syrup urine disease (MSUD), tyrosinosis, and alkaptonuria.
2. Aminoacidurias characterized by elevated amino acid urine levels with normal blood levels: *transport diseases* with a defect in the kidney tubule (eg, cystinuria), and "no-threshold" aminoaciduria in which the kidney has no mechanism for reabsorbing the amino acid involved (eg, homocystinuria).

PKU, a disease characterized by mental deficiency, is associated with the presence of phenylpyruvic acid in the urine and elevated serum phenylalanine levels due to a hereditary (autosomal recessive) deficiency of hepatic phenylalanine hydroxylase, which converts phenylalanine to tyrosine. The availability of treatment through dietary intake is predicated upon early detection. Many states have passed legislation for mass-screening for PKU in all infants. The Guthrie test is performed by placing filter paper discs impregnated with serum or blood on the surface of an agar culture medium containing β-(2-thienyl)alanine at a concentration sufficient to inhibit the growth of *Bacillus subtilis*. Phenylalanine will reverse this inhibition, and the Bacterial Inhibition Assay (BIA) is a direct measure of this amino acid. Serum phenylalanine determinations also can be performed by estimating the fluorescence of a complex with ninhydrin and copper in the presence of L-leucyl-L-alanine.

MSUD is characterized by the odor of the urine and is rapidly fatal to infants. It is associated with a deficiency in the oxidative decarboxylation of α-keto acids leading to an accumulation of both the keto and amino acids in the blood and urine (valine, leucine, isoleucine). TLC and BIA assays can be used to detect MSUD.

Alkaptonuria is a rare, hereditary disease in which homogentisic acid cannot be metabolized further due to a lack of homogentisic acid oxidase. This causes homogentisic acid-uria, ochronosis, and arthritis.

In *Hartnup disease*, indole and tryptophane appear in the urine due to defective renal and intestinal absorption of tryptophane. Tryptophane is an intermediary metabolite in the synthesis of *serotonin* (5-hydroxytryptamine) and 5-hydroxyindole acetic acid (HIAA). Excessive production of *serotonin* and the presence of its *HIAA metabolite* in the urine are associated with metastatic carcinoid tumors. HIAA is measured after removal of interfering keto acids with dinitrophenylhydrazine, extraction, and estimation with nitrosonaphthol reagent.

Routine screening tests for congenital metabolic defects and the substance under test in the newborn include PKU (phenylalanine), MSUD (leucine), tyrosinemia (tyrosine), homocystinuria (methionine), histidinemia (histidine), valinemia (valine), galactosemia (galactose or galactose uridyltransferase), orotic aciduria (orotidine-1-phosphate decarboxylase), arginosuccinuria (arginosuccinic lyase), hereditary angioneurotic edema (C^1-1-esterase inhibitor), and sickle-cell disease (hemoglobin S).

The analyses for these substances are based on BIA, metabolite bacterial inhibition assay (MIA), enzyme auxotroph bacterial assay (ENZ-Aux), fluorescent spot tests or TLC, and electrophoresis.
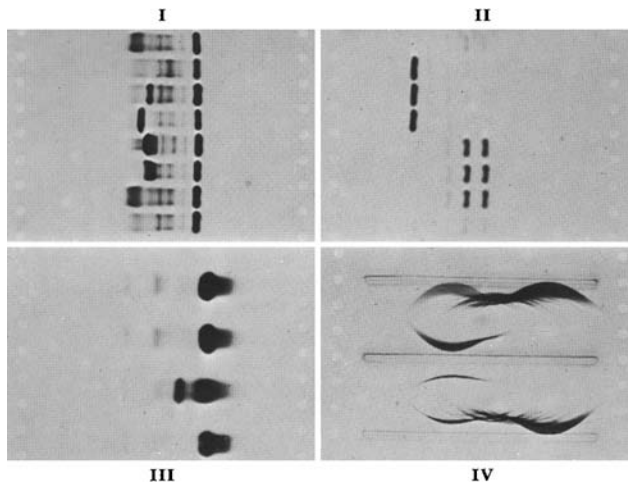
PROTEINS

The *plasma proteins* (albumins, globulins, and fibrinogen) are involved in nutrition, electrolyte and acid–base balance, transport mechanisms, coagulation, immunity, and enzymatic action. *Total plasma proteins* may be determined by Kjeldahl, Nesslerization, specific ion pair (bromcresol green dye plus albumin), or biuret procedures. The last technique is based on the reaction of —CONH— groups joined by carbon or nitrogen linkages in protein with alkaline copper sulfate to yield the biuret complex that can be estimated colorimetrically. Total protein also can be estimated by specific gravity, or refractometric or UV spectrometric methods. These methods are subject to large errors in the presence of a pathology involving increased glucose, lipid, urea, or abnormal protein concentrations.

The *albumin-globulin* (A/G) *ratio* is determined by the biuret method after precipitation of the globulins with a sodium sulfate–sulfite reagent. The normal range is 5.5 to 8.0 g/dL total protein with an A/G ratio of 1.4 to 2.4. Changes in total protein and A/G ratio occur in kidney and liver disease, hemorrhage, dehydration, rheumatoid arthritis, and multiple myeloma. Gastrointestinal albumin loss, as seen in GI bleeding, ulcerative colitis, sprue, and enteritis, can be detected by monitoring fecal radioactivity after intravenous injection of ^{51}Cr-human serum albumin.

The physiochemical properties of the plasma proteins—(mol wt 68,000 to 300,000 daltons) and isoelectric point (pH of minimum solubility and ionic neutrality)—provide the basis for the electrophoretic separation of plasma proteins (Fig 32-4). The plasma sample is spotted on a paper or cellulose acetate strip, or in a polyacrylamide gel (disc or gel electrophoresis) at pH 8.6. At this pH the proteins are electroanionic and, under the influence of electric current, will migrate to the anode at a rate dependent on their isoelectric point and, in the case of cellulose acetate or gel electrophoresis, their molecular size. The strips are then stained with a protein dye (bromophenol blue, Amido black, or Ponceau S), and the concentrations of the various proteins are estimated by densiometric scanning.

The normal ranges for the major proteins are (in g/dL): albumin 3.8 to 5.0; total globulin, 2.0 to 3.9; α_1-globulin, 0.1 to 0.5; α_2-globulin, 0.5 to 0.9; β-globulin; 0.5 to 1.2; γ-globulin, 0.7 to 1.6.

Ordinary electrophoresis does not identify the subgroups of *immunoglobulins*, IgA, IgM, IgG, and IgE. This is accomplished by immunoelectrophoresis, a process involving electrophoresis and immunodiffusion. The sample is electrophorized in an agar gel (zone electrophoresis) and then antiserum to the specific Ig or to total globulins is placed in a trough aligned parallel to the axis of the original electrophoresis. The serum proteins and

Figure 32-4. Electrophoretic separation of serum proteins (I), isoenzymes (II), hemoglobins (III), and immunoelectrophoresis of plasma protein (IV) (courtesy, Spinco).

antisera diffuse toward each other and form precipitin (antigen-antibody complex) lines. Ordinary cellulose acetate or gel electrophoresis will permit the recognition of diffuse, polyclonal elevation of serum immunoglobulins seen in chronic infections, isolated M-protein peaks of macroglobulinemia and multiple myeloma, and absent gamma component in a hypogammaglobulinemia or agammaglobulinemia. Immunoelectrophoresis will indicate specific Ig abnormalities or, by noting the presence of any displacement, bowing or broadening of the precipitin band will aid in the diagnosis of the paraimmunoglobulin monoclonal diseases such as multiple myeloma, macroglobulinemia, or chronic lymphatic leukemia.

Radial immunodiffusion is a simple process that also can be used for quantitation of IgA, IgM, and IgG.[16] It is performed by incorporating the antibody in an agar gel and then introducing the antigen or test sera into wells punched in the agar. The antigen diffuses radially out of the well into the surrounding gel media, and a visible precipitin line forms where the antigen and antibody have reacted. Quantitation of IgA, IgM, and IgG aids in the diagnosis and differentiation of collagen diseases, chronic infections, and liver disease. IgE is best quantitated by immunoelectrophoresis or RIA.

Nephelometric techniques detect immunological constituents by measuring the light-scattering properties of various antigen-antibody complexes in a test solution. The Hyland system measures the amount of laser-beam deflection at an angle by employing a photomultiplier tube that is sensitive in the red region of the spectrum. Results are calculated by an electronic-screening system and read in percent relative light-scatter on a digital readout. Automated electrophoresis instrumentation offers computer-controlled sample application, staining options, densitometry, and pattern interpretation for serum proteins and isoenzymes.

ENZYMES

Enzymes are proteins whose biological function is the catalysis of chemical reactions in living systems. Enzymes combine with the substances on which they act (substrates) to form an intermediate enzyme-substrate complex, which is then converted to a reaction product and liberated enzyme that continues its catalytic function. Enzymes are highly specific; a few exhibit absolute specificity and catalyze only one particular reaction, whereas others are specific for a particular type of chemical bond, functional group, or stereoisomeric structure.

Most serum enzymes of clinical significance are intracellular in origin and are elevated in hyperactivity disease, malignancy,

or injury to cardiac, hepatic, pancreatic, muscle, bone, and tissue. As the specific tissue involved will determine the type of enzyme that will be elevated, such determinations are valuable diagnostic tools in the differentiation of various pathological states.

Enzymes are named and classified according to the type of reaction that they catalyze and to their substrate specificities. Enzyme activity usually is expressed in International Units (IU) where 1 unit (U) is that amount of the enzyme that will catalyze the transformation of 1 μmole of substrate/min at a specific temperature, pH, and substrate-concentration conditions.

Transferases are enzymes that catalyze the transfer of amino or phosphate groups from one compound to another. Aspartate aminotransferase (AST) and alanine aminotransferase (ALT) are important in clinical diagnosis. These enzymes catalyze the transfer of the amino group from glutamic acid to keto acids (oxaloacetic or pyruvic) to form aspartic and α-ketoglutaric acids with AST (aspartate aminotransferase) and alanine and α-ketoglutaric acid with ALT (alanine aminotransferase).

Colorimetric methods are based on an estimation of the reaction products (oxaloacetic or pyruvic acid) with dinitrophenylhydrazine, or substrate (α-ketoglutaric acid) by coupling with 6-benzamido-4-methoxy-*m*-toluidinediazonium chloride.

Spectrometric methods are based on the reaction of the product pyruvate with lactic dehydrogenase and NADH, or of oxaloacetate with malic dehydrogenase and NADH. The rate of NADH utilization is measured by the decrease in absorbance at 340 or 360 nm and is directly proportional to transaminase activity.

Normal AST and ALT levels are < 40 U/L. AST is present in large amounts in liver, cardiac, and skeletal muscle, whereas ALT is found primarily in liver tissue. AST is elevated in myocardial infarction and Duchenne muscular dystrophy; AST and ALT are increased in liver disease, acute toxic or viral hepatitis, infectious mononucleosis, obstructive jaundice, and hepatic cirrhosis.

Creatine kinase (CK) is a transferase found in muscle and brain tissue. It catalyzes the transfer of phosphate groups from creatine phosphate to ADP to form ATP. Activated CK activity is measured by following the increase of ATP in the creatinine phosphate-ADP reaction in the presence of glutathione or cysteine thiol activators. The ATP can be measured by the fluorometric determination of light emitted by luciferinase conversion of luciferin to adenyl-oxyluciferin in the presence of ATP. Normal serum levels are < 50 U/L; it is elevated in myocardial infarction and Duchenne muscular dystrophy, but remains at normal levels in liver disease.

Ornithine transcarbamylase (OTC) in serum is the only enzyme of the urea cycle that has been used in the clinical investigation of liver disease. It catalyzes the conversion of ornithine to citrulline. The normal serum value is 0 to 0.4 U/L.

Oxidoreductases or *dehydrogenases* are enzymes that catalyze hydrogen transfer in cellular oxidation processes. *Lactic (LDH), α-hydroxybutyric (HBDH), malic (MDH), glutamic (GLDH), isocitric (ICDH),* and *sorbitol (SDH) dehydrogenases* are of diagnostic importance in myocardial and liver disease.

LDH catalyzes the reversible conversion of pyruvic to lactic acid in the presence of NADH. The activity may be estimated colorimetrically by forming the pyruvic acid hydrazone with 2,4-dinitrophenylhydrazine; spectrometric or fluorometric estimation of NADH in this reaction also is used to estimate enzyme activity. The normal serum LDH value is < 200 U/L (pyruvic → lactic) and < 50 U/L (lactate → pyruvate). LDH is increased to a much greater extent and for a more prolonged period than AST or CK in myocardial infarction; it also is increased to varying degrees in certain types of hepatic disease, disseminated malignancies, pernicious anemia, and muscular dystrophy.

Recent advances in protein chemistry and technical methodology have led to fractionation of enzymes, previously thought to be homogeneous, into heterogeneous moieties. These multiple-molecular forms of enzymes (*isoenzymes*) have similar substrate specificity but different biophysical properties. LDH,

MDH, CK, phosphatases, and leucine aminopeptidase exist in isoenzyme forms.

CK isoenzymes are important in the early detection of myocardial damage. Two CK molecular subunits, M and B, produce three isoenzymes: CK-MM found primarily in skeletal muscles, CK-MB in the myocardium, and CK-BB primarily from the brain. After acute myocardial infarction (MI), CK-MB appears in the serum in approximately 4 to 6 hr, reaches peak activity at 18 to 24 hr, and may disappear within 72 hr. Diagnostic testing of MI includes CK and LDH isoenzymes. Early detection of CK-MB allows the management of myocardial infarcts with agents such as streptokinase or tissue plasminogen activator (tPA). The methods of assessment include electrophoresis, column chromatography, and immunoinhibition.

Serum contains five LDH isoenzymes, each a tetramer composed of one or two monomers. LDH 1 and 2 are found in preponderance in heart, kidney, and RBC, whereas liver and skeletal muscle largely contain LDH 4 and 5. Intermediate forms prevail in lymphatic tissues and many malignancies. The fractionation of LDH isoenzymes is important in the differential diagnosis of cardiac, muscle, and liver disease. It can be accomplished with DEAE-cellulose chromatography, electrophoresis, sulfite, or urea inhibition of specific isoenzymes, thermal stability, and substrate-concentration requirements.

HBDH reduces α-ketobutyric acid to α-hydroxybutyric acid in the presence of NADH; estimation of the α-keto acid via hydrazone formation or NADH is the basis of activity measurements. The normal serum HBD level is <140 U/L; it is elevated in myocardial infarction. LDH 1 is high in HBDH activity. The ratio of total LDH/HBDH often is used in place of LDH isoenzyme determination. Ratios > 0.8 are seen in myocardial infarction and <0.6 in acute liver damage.

MDH and *SDH*, in the presence of NAD, catalyze the conversion of malate or sorbitol to oxaloacetate or fructose, respectively. They are of diagnostic value in MI (MDH > 48 U/L) and acute liver injury (SDH > 96 U/L).

ICDH oxidizes isocitrate, in the presence of NADP or NAD, to α-ketoglutarate; it is elevated (>5.0 U/L) in acute hepatitis.

Hydrolases are enzymes that catalyze the addition of the elements of water across the bond that is cleaved.

Amylases, lipases, phosphatases, 5′-nucleotidase, γ-glutamyl transferase, and *leucine aminopeptidase* are specific examples of clinically important hydrolases.

Salivary and pancreatic *amylases* hydrolyze the substrate starch to maltose and dextrins. Amylase activity can be measured by procedures based on the loss in certain properties of starch as it is hydrolyzed (*amyloclastic*), or by the generation of reducing substances (*saccharogenic*). The amyloclastic methods use the decrease in viscosity and turbidity of hydrolyzed water-soluble starch substrates, or the reaction of starch with iodine as the method of estimation. A newer procedure uses an improved substrate, ethylidene-G$_7$PNP, which prevents undesired hydrolysis of the substrate by α-glucosidase. This results in greater accuracy in amylase testing. Normal serum level using this methodology is ≤88 U/L; elevations are noted in acute pancreatitis, acute abdominal conditions (perforated peptic ulcer, common bile-duct obstruction), and salivary gland disease.

Lipases catalyze the conversion of triglycerides to glycerol and fatty acids. Classic clinical determination was based on the titrimetric analysis of fatty acids liberated from an emulsified olive oil substrate, a slow, tedious methodology requiring several hours of incubation. Modern methods are based on the hydrolysis by pancreatic lipase of 1,2-diglyceride to 2-monoglyceride and fatty acid. The 2-monoglyceride then is measured by coupled-enzyme reactions catalyzed by monoglyceride lipase, glycerol kinase, glycerolphosphate oxidase, and peroxidase. This assay is simple to perform and can be adapted easily to automated analyses. The measurement of serum lipase is used widely for the diagnosis of acute pancreatitis, in which a 10-fold increase above the upper reference limit (60 U/L) is sug-

gestive of pancreatitis, pancreatic injury, or inflammation of organs contiguous to the pancreas.

Phosphatases catalyze the hydrolysis of orthophosphoric acid esters and are classified according to the pH of optimal activity into alkaline or acid phosphatases. Activity (alkaline, pH 8 to 10; acid, pH 4 to 6) is measured with phenyl phosphate, glycerophosphate, *p*-nitrophenyl phosphate, or thymolphthalein monophosphate substrates. With the latter two chromogenic substrates, the amount of *p*-nitrophenol or thymolphthalein liberated by phosphatase hydrolysis is estimated colorimetrically in an alkaline medium. With a glycerophosphate or phenyl phosphate substrate, the liberated phosphorus is determined by molybdenum blue formation with phosphomolybdic-phosphotungstic acids; phenol also may be estimated with 4-aminoantipyrine or Folin-Ciocalteau reagent.

Acid phosphatase activity may be differentiated by the use of inhibitors in the assay mixture; formaldehyde has no effect on acid phosphatase of prostatic origin, but it inhibits other acid phosphatases, whereas tartrate is a selective inhibitor of the prostatic enzyme. *Acid phosphatase* is of a primary diagnostic value in metastatic carcinoma of the prostate. Normal values for *alkaline phosphatase* activity depend on the substrate used; elevations in osteomalacia and in bone tumors depend on the degree of osteolytic or osteoblastic activity. The enzyme (isoenzyme) also is elevated in obstructive jaundice, and bone and liver disease.

The enzyme *5′-nucleotidase* is an alkaline phosphomonoesterase that hydrolyzes nucleotides with a phosphate radical attached to the 5′-position of the pentose (eg, adenosine monophosphate). The normal serum value is 17 U/L; it is elevated in hepatic disease.

Leucine aminopeptidase (LAP) is an exopeptidase that hydrolyzes the peptide bond adjacent to a free amino group. It liberates amino acids from the *N*-terminal group of proteins and polypeptides in which the free amino group is an L-leucine residue. Activity is determined by spectrophotometric estimation following hydrolysis of the amide bond of a leucin-amide substrate at 238 nm. Clinical estimations usually are performed on synthetic substrates, and as there is no correlation between cleavage of leucinamide and these substrates, the LAP-like activity is designated *leucine arylamidase*. A fluorometric determination of naphthylamine liberated from a leucyl-β-naphthylamide substrate or colorimetric determination of *p*-nitroaniline liberated from leucine-*p*-nitroanilide substrate also has been used. The normal value is 8 to 22 U/L; it is elevated in the last trimester of pregnancy, hepatobiliary disease, and pancreatic carcinoma.

Serum *γ-glutamyl transferase* (γGT) is increased in diseases of the liver, bile ducts, and pancreas. Together with alkaline phosphatase, LAP, and 5′-nucleotidase, γGT usually is tested in the group of cholestasis-indicating enzymes. The assay is based on the hydrolysis of γ-glutamyl-*p*-nitroanilide.

Serum lysozyme (muramidase) activity is increased in certain types of leukemia. Serum arginase, an enzyme that hydrolyzes arginine to ornithine and urea, and serum guanase are sensitive indicators of hepatic necrosis.

Lyases are enzymes which split C—C bonds without group transfer. *Aldolase* is a glycolic lyase that catalyzes the reversible splitting of fructose 1,6-diphosphate to form dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. In the estimation of activity, the triose phosphate reaction products are hydrolyzed with alkali and the resultant trioses are reacted with 2,4-dinitrophenylhydrazine to form chromogenic hydrazones for colorimetric analysis. A spectrophotometric estimation is made by coupling the aldolase reaction products with a dehydrogenase acting on one of the triose phosphates and measuring concomitant changes in NADH. The normal value is < 8 U/L; it is elevated in muscular dystrophy, polymyositis, and acute hepatitis.

The significance of serum-enzyme changes in hepatitis is seen in Figure 32-5 and enzyme activity following myocardial infarction in Figure 32-6.
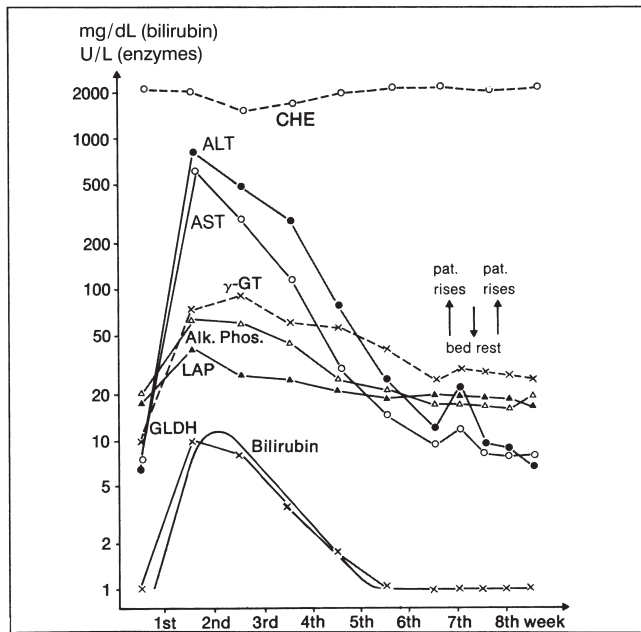
Figure 32-5. Typical course of alterations in serum enzyme activity in acute viral hepatitis. (Adapted from Schmidt E, Schmidt FW *Med Welt* 1970; 21:805.)

Figure 32-6. Serum enzymes following myocardial infarction, AST, CK, LDH, and HBD are compared.
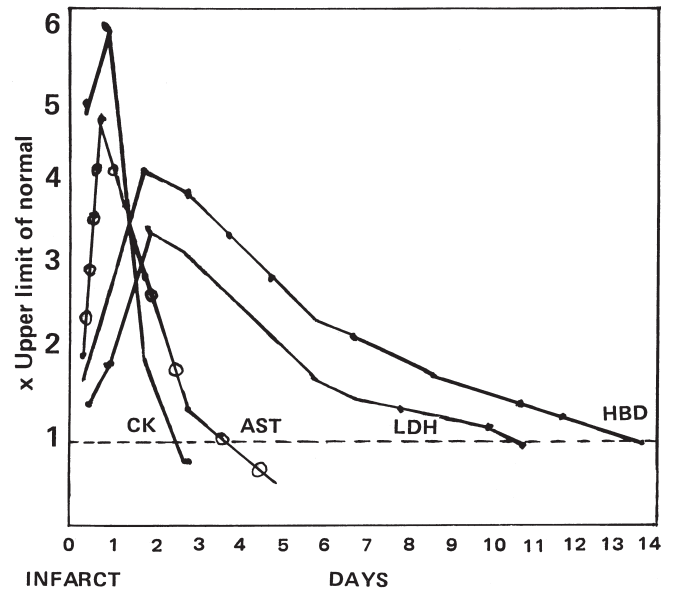
LIPIDS

The major classes of blood lipids are *fatty acids, cholesterol, triglycerides, phospholipids,* and *lipoproteins.* Hyperlipidemia is not a single aberration, and there are a number of different hyperlipidemic states. Lipid-profile tests include measurements of cholesterol, triglyceride, phospholipids, and determination of lipoprotein phenotypes.

Cholesterol, a sterol molecule, is an essential substance in steroid-hormone synthesis by the adrenal cortex and bile acid production in the liver. It exists in blood as the free sterol and as cholesterol esters of fatty acids.

In the determination of *total cholesterol*, the serum is extracted with an alcohol–ether mixture and the cholesterol estimated colorimetrically after reaction with acetic anhydride–sulfuric acid reagent (Liebermann-Burchard reaction). The precipitation of free cholesterol with digitonin will differentiate free from esterified cholesterol. Chromatographic separation of cholesterol from its esters on alumina, silicic acid, or magnesium silicate columns with organic solvents also has been used.

Gas chromatographic procedures have resulted in the separation and quantitation of cholesterol, its metabolites, and precursors; this is a type of partition chromatography in which a volatilized sample is partitioned between a liquid stationary phase and a mobile gas phase. The normal-adult total-serum-cholesterol level is 150 to 270 mg/dL; it is increased in hyperlipemia and specifically in hyper-β-lipoproteinemia, nephrosis, diabetes mellitus, and myxedema, and decreased in hyperthyroidism and hepatic disease. Free cholesterol comprises 20% to 40% and the ester fraction 60% to 80% of the total serum cholesterol.

Phospholipids are "compound" or "heterolipids" that contain phosphorus, a nitrogen base and a long-chain fatty acid. Lecithin (phosphatidylcholines) and cephalin (phosphatidylethanolamine or serine) are the principal plasma phospholipids, which normally comprise one-third of the total plasma lipids. They usually are bound to lipoproteins. These serum lipids are extracted into an alcohol–ether mixture, digested with sulfuric acid-hydrogen peroxide and the liberated phosphorus determined by colorimetric techniques. The normal lipid phosphorus is 6 to 11 mg/dL; about one-half is lecithin. The average ratio of cholesterol to lipid phosphorus when cholesterol is normal is 21. Phospholipid

changes usually are associated with cholesterol changes and are of interest in coronary artery and liver diseases and the hyperlipoproteinemias.

Sphingolipids differ from lecithin and cephalin. They are phosphate esters of sphingosine bound to choline or ethanolamine and primarily are found in brain tissue (eg, sphingomyelin, galactolipin). The *ratio of lecithin to sphingomyelin* (L/S) in amniotic fluid or resuscitated amniotic fluid from the oral cavity of the newborn is an accurate assessment of fetal maturity and the respiratory-distress syndrome. Changes in phospholipid biosynthesis during gestation reflect the aging of the fetal lung, as the L/S ratio normally increases.

Tay-Sachs disease is a lipid-storage disease in which the central nervous system degenerates because of the progressive intraneuronal accumulation of excess amounts of the sphingolipid ganglioside GM_2. The accumulation of GM_2 in Tay-Sachs disease has been shown to be caused by a lack of the enzyme hexosaminidase A. Therefore, the measurement of serum, WBC, or amniotic fluid *hexosaminidase A* is important in evaluating carriers and in diagnosing Tay-Sachs disease in the fetus.

Both hexosaminidase A (heat-labile) and hexosaminidase B (heat-stable) can catalyze the conversion of 4-methylumbelliferyl-*N*-acetylgalactosamine (a synthetic substrate) to *N*-acetylgalactosamine and 4-methylumbelliferone. The cleavage product, 4-methylumbelliferone, fluoresces under ultraviolet radiation, and the intensity of the fluorescence is a measure of the activity of the enzyme. In noncarriers, 50% to 75% of the total hexosaminidase activity is heat-labile (hexosaminidase A), and in carriers 20% to 45% of the total hexosaminidase activity is heat-labile.

The blood fatty acids occur in esterified (EFA) and nonesterified (NEFA) forms. *Triglyceride* determinations are of value in differentiating the hyperlipidemic states, that is, essential (diet-induced) hypertriglyceridemia from familial hypocholesterolemia with or without triglyceridemia. After the preliminary separation from phospholipids, triglycerides most often are determined in terms of their glycerol moiety. The glycerol released by saponification is oxidized to formaldehyde and the latter determined by fluorometric or colorimetric procedures. Triglycerides also can be determined by coupling the glycerol liberated from lipase/α-chymotrypsin treatment of serum with a glycerol kinase–pyruvate kinase–LDH system and spectrometric estima-

tion of NADH. Normal triglyceride levels are 110 to 140 mg/dL. An increase in triglycerides will produce a milky appearance in serum (lipemic). EFA analyses are based also on the reaction of alkaline hydroxylamine with esters of fatty acids to form hydroxamic acids that produce a red color with ferric chloride.

Gas chromatographic procedures have been used to quantitate the various *fatty acids,* that is, palmitic, stearic, oleic, linoleic, and linolenic acids. Mono-, di-, and triglycerides also can be separated into classes and quantitated by column or thin-layer chromatography, and infrared spectrometry. The total fatty acids of plasma range from 200 to 450 mg/dL in the fasting state; they are derived from glycerides, cholesterol esters, and phospholipids.

All the lipids in plasma circulate in combination with protein. The free fatty acids are bound to albumin, and the lipids aggregate with other proteins to form *lipoproteins*. Electrophoresis and ultracentrifugation are the principal methods used to separate and identify lipoprotein families. *Chylomicrons* ($S_f > 400$), *pre-β-lipoproteins* (S_f 20–400), *β-lipoproteins* (S_f 0–20), and *α-lipoproteins* are the four major classes in order of increasing density and migration on cellulose acetate electrophoresis. Chylomicrons are representative primarily of dietary or exogenous triglycerides, pre-β-lipoproteins of endogenous glycerides, β-lipoproteins of cholesterol and its esters, and α-lipoproteins of cholesterol and phospholipids. Abnormal lipoproteins that may appear in plasma include floating β-lipoproteins, lipoprotein X, and complexes of normal lipoproteins with IgA and IgG myeloma proteins (autoimmune hyperlipoproteinemia). Age, sex, diet, fasting, posture changes, and trauma can alter the lipid profile.

The lipoprotein classes usually are separated by paper, agarose, or cellulose acetate electrophoresis. The strips are stained with fat-soluble dyes (Sudan Black or Oil Red O) and quantitated by densiometric scanning. Primary hyperlipoproteinemias are classified into normal and five abnormal types based on cholesterol and triglyceride levels and lipoprotein analysis. Hyperchylomicronemia (type I), hyper-β-lipoproteinemia (type II), broad β-band (type III), hyper-pre-β-lipoproteinemia (type IV), and hyper-pre-β-lipoproteinemia and chylomicronemia (type V) are the major classes. Carbohydrate and fat-tolerance studies, post-heparin lipase activity, and clinical symptomatology also are integrated into the diagnosis of the various subclasses. The presence or predisposition to coronary artery disease and other disease states is associated with the various types.[17]

STEROIDS AND OTHER HORMONES

The steroids possess a common structure, the perhydrocyclopentanophenanthrene nucleus, and include cholesterol, bile acids, androgens, and the adrenocortical, adrenomedullary, estrogenic, and progestational hormones.

Androsterone, dehydroepiandrosterone, etiocholan-3α-ol-17-one, 11-ketoandrosterone, 11-ketoetiocholanolone, 11β-hydroxyandrosterone, and 11β-hydroxyetiocholanolone are the principal *urinary 17-ketosteroids (17KS).* These androgenic hormones are derived from the adrenal and, in males, testicular function. The principal urinary steroid metabolites in this group of androgens are found both in the free form, and as conjugates of glucuronides, sulfates, or acetates. Their determination in urine involves the acid hydrolysis of the conjugates, extraction with organic solvent, reaction with alkaline *m*-dinitrobenzene (Zimmerman reaction), and colorimetric estimation of the chromogen. The individual 17KS can be separated by TLC prior to analysis to obtain further information on the individual steroids. The normal adult urine values are: male, 9 to 24 mg/day; female, 5 to 17 mg/day. Decreased excretion is seen in hypoactive disease of the pituitary, gonads, and adrenals. Increased excretion is seen in hyperplasia, cancer, or tumors of the adrenals.

Testosterone is the most potent androgen in blood. The measurement of urinary or serum testosterone is useful in distinguishing normal and hypogonadal males and in treating hirsutism in the female. This hormone is determined by gas chromatography, competitive protein-binding, isotope dilution, or RIA procedures. Normal serum testosterone is 0.2 to 1.1 µg/dL in the male and <0.1 µg/dL in the female.

The natural *estrogenic hormones* are estradiol, estrone, and estriol, produced in the gonads, adrenals, and placenta. The relative amounts of the three estrogens rise and fall concomitantly during the menstrual cycle. Maternal urinary total-estrogen excretion, especially estriol, is an indirect index of the integrity and viability of the fetoplacental unit. Analysis involves acid or glucuronidase–arylsulfatase hydrolysis of the conjugates, removal of urinary glucose if present, extraction, and colorimetric or fluorometric analysis. In the determination, after acid hydrolysis and ether extraction of the urine, the estrogens are methylated with dimethyl sulfate and chromatographically separated prior to reaction with phenolsulfuric acid to yield a red chromogen for colorimetric analysis. The normal estrogen output is 4 to 60 µg/24 hr in the female and up to 25 µg in the male. Estrogen deficiency can be related to ovarian failure and pituitary deficiency.

Progesterone is a progestational hormone that is secreted by the corpus luteum of the ovary and also by the adrenal cortex. Serum progesterone determination is of value in the detection of ovulation and is a measure of the secretory activity of the placenta during pregnancy. Progesterone is determined in serum by RIA, double-isotope derivatization, gas-liquid chromatography, or competitive protein-binding techniques. Normal menstrual-cycle serum progesterone levels vary between 0 and 1.6 µg/dL.

Pregnanediol is the principal metabolite of progesterone. The urinary determination of pregnanediol excretion is an indirect index of progesterone levels but is subject to variation due to individual differences in hepatic metabolism of this hormone and is not representative of total endogenous progesterone production.

Adrenal cortex steroids include glucocorticoids, androgens, estrogens, progesterone, and mineralocorticoids. Glucocorticoids can be determined as plasma cortisol (plasma 17-OH corticosteroids), urinary-free (unconjugated cortisol), or total-urinary *17-OH corticosteroids*. The latter are determined in urine as *17-ketogenic steroids* (17KGS). The 17KS in urine are reduced with borohydride to alcohols; the 17-OH steroids are oxidized with sodium bismuthate or periodate to 17KS and quantitated by the alkaline dinitrobenzene method. The 17-OH steroids can be quantitated directly by the phenylhydrazine-sulfuric acid reaction after hydrolysis of glucuronide conjugates and chromatographic purification. The 17-OH steroid analysis only determines compounds with the dihydroxyacetone side chain, such as tetrahydrocortisol or tetrahydrocortisone; the 17KGS analysis includes the 17-OH-corticosteroids with the dihydroxyacetone side chain and the pregnanetriol type of compound. Normal 17KGS daily urinary excretion is 5 to 23 mg in the male and 3 to 15 mg in the female. They are reduced significantly in myxedema and adrenal or anterior pituitary insufficiency. Plasma cortisol usually is measured by fluorometric or gas chromatographic procedures.

Aldosterone is the most active member of the mineralocorticoid group. The determination of urinary aldosterone is of value in differentiating benign essential hypertension from primary aldosteronism (Conn's syndrome), which is caused by an adrenal adenoma and is accompanied by hypertension. A double-isotope derivatization technique is used. Urinary aldosterone is acetylated with ³H-acetic anhydride; aldosterone-¹⁴C-diacetate standard is added early in the procedure. The ³H/¹⁴C specific activity of the final product is measured after chromatographic purification and is a direct measurement of aldosterone. The normal aldosterone levels of about 10 µg/day are elevated in Conn's disease and usually are associated with low serum potassium, sodium retention, and low-concentration alkaline urine.

The anterior pituitary secretes three substances (*gonadotropins*) that regulate gonadal activity: *follicle-stimulating hormone (FSH), luteinizing hormone (LH)* or *interstitial cell*

hormone *(ICSH),* and *luteotropin (LTH).* The gonadotropins are glycoproteins. Bioassay methods can be used to determine gonadotrophic activity. After fractionation and isolation, the urine extract is assayed in test animals as to the follicular growth of the ovaries in hypophysectomized animals or - increase in testicular, ovarian, or uterine weight in various– animal models. RIA techniques have been developed for these gonadotropins and represent the most sensitive and precise measurement method, although nonisotopic enzyme immunoassays (EIA) rapidly are becoming popular due to the increased costs of isotope disposal and their analytical performance equaling or surpassing RIA procedures.

Analysis of serum or urinary *placental lactogen (HPL)* and *chorionic gonadotropin (HCG),* a placental-derived protein hormone, is useful in the diagnosis of threatened abortion, hydatiform mole, and choriocarcinoma. HCG, pregnanediol, and progesterone as well as total and fractionated estrogens are useful in testing for pregnancy. HCG and HPL readily are measured by RIA or EIA, and low values are seen in threatened abortion and intrauterine fetal death.

The increase in HCG in the serum or urine of the pregnant female is the basis of a routine *pregnancy test*. Test components consist of an antigen in the form of HCG latex particles and an HCG antiserum. When antiserum is mixed with urine containing a detectable level of HCG, it is neutralized and no agglutination of latex-antigen particles occurs (*agglutination inhibition test*). The commercial application of the HCG assay gives laboratories a rapid, accurate pregnancy test by taking advantage of monoclonal antibody specificity and sensitivity. A monoclonal slide procedure on urine, Duoclon (*Organon Diagnostics*), uses two different monoclonal antibodies, one against HCG and one against the HCG_B subunit for maximum specificity. Agglutination indicates a positive test with a sensitivity level of 500 miU HCG/mL, detecting pregnancy a few days after conception.

Human growth hormone and insulin are proteins that are of diagnostic value in growth-rate studies and diabetes. They are best quantitated by RIA.

Epinephrine and *norepinephrine* are biologically active catecholamines derived from the adrenal medulla and sympathetic nerve endings. Catecholamines are measured in the blood and urine after fractionation on alumina or ion-exchange columns, oxidation at pH 3.5 or 6, and subsequent fluorometric analysis. Urine catecholamines are increased to > 350 µg/24 hour in adrenal medullary tissue tumors (pheochromocytoma). The normal plasma level is 2.1 to 6.5 µg/L with about 80% as norepinephrine.

Vanillylmandelic acid (VMA) is the urine metabolite of these two catecholamines. Its quantity in urine reflects the endogenous secretion of catecholamines. VMA can be determined colorimetrically, after extraction of the urine with ethyl acetate, and diazotization with *p*-nitroaniline and ethanolamine in the presence of carbonate ion. VMA also can be measured spectrometrically following periodate oxidation to vanillin and solvent extraction. The normal output is 0 to 12 mg/24 hour.

Homovanillic acid (HVA) is not a metabolite of epinephrine or norepinephrine, but is produced from a common precursor, dopamine. Elevated HVA excretion is diagnostic in cases of neuroblastoma.

The biosynthesis of *serotonin* (5-hydroxytryptamine) and urinary excretion of its metabolite, 5-hydroxyindoleacetic acid (5-HIAA), are increased in argentaffine tumors. These have a very large capacity to metabolize tryptophane stores to serotonin. Urinary 5-HIAA increases from 1 to 7 mg/24 hr to as much as 1 g/24 hr in this type of tumor.

Bilirubin, a tetrapyrrole that is derived from senescent red-cell degradation, normally occurs in low concentration in the blood. In bile, it is present as the water-soluble conjugated acyldiglucuronide. In blood, bilirubin is bound tightly to plasma albumin. The reduction of bilirubin in the intestine yields urobilinogen, which is, in turn, oxidized to a brown pigment—urobilin.

Serum bilirubin is determined by coupling with diazotized sulfanilic acid to form azobilirubin for colorimetric analysis. The *direct* or *conjugated bilirubin* test is performed in aqueous media; the *indirect* or *free bilirubin* analysis is performed in methanol or caffeine-sodium benzoate solution. Normal values in serum are: direct, 0 to 0.3 mg/dL; total, 0 to 1.5 mg/dL.

Clinical jaundice is a yellowing of the tissues associated with hyperbilirubinemia; in hemolytic disease of the newborn due to Rh and ABO incompatibilities, indirect serum bilirubin is elevated, whereas acute hepatitis results in increases in the direct type.

ELECTROLYTES

The normal plasma electrolyte level is 154 mEq/L of cations and 154 mEq/L of anions. The osmotic effects of chloride, bicarbonate, sodium, and potassium are important in the maintenance of normal muscle contraction and water distribution between cells, plasma, and interstitial fluid.

Flame photometry, atomic-absorption spectrometry, neutron-activation analysis, x-ray fluorescence, ion-specific electrodes, and colorimetric techniques are used in the identification and determination of cations or anions in biological fluids. Advances in technology have developed multiphase systems capable of measuring not only sodium and potassium but also chloride, carbon dioxide, and calcium simultaneously.

Sodium and *potassium* serum concentrations are readily measured by flame photometry or highly sensitive and specific atomic-absorption spectrometry. The latter technique is similar to emission-flame photometry, except that it measures energy as it is absorbed by atoms rather than as it is emitted by atoms. Both techniques are based on the characteristic absorption or emission wavelengths of the cations. Ion-specific electrodes also are used for Na^+ and K^+ determinations, eliminating the use of a flame or combustible gas, and this can be performed on whole blood, plasma, or serum.

Chloride levels in serum or urine are determined by titration with acid mercuric nitrate solution in the presence of *s*-diphenylcarbazone indicator. They also may be determined potentiometrically with a silver–silver chloride pH electrode assembly. The normal serum values are 135 to 155 mEq Na/L, 3.9 to 5.6 mEq K/L, and 95 to 106 mEq Cl/L; urine levels are 150 to 197 mEq Na/day, 20 to 64 mEq K/day, and 180 to 270 mEq Cl/day.

Serum sodium, potassium, chloride, and *bicarbonate* determinations are useful indicators in adrenal cortical insufficiency, renal and cardiac failure, anuria, dehydration, alimentary tract diseases associated with diarrhea and vomiting, and increased renal electrolyte excretion (diuretic therapy).

The determination of excess *chloride* (>50 mEq/L) in the perspiration of patients with pancreatic *cystic fibrosis* is an accurate diagnostic tool. Perspiration is stimulated by placing the patient's hand in a plastic bag for 15 to 20 min or, preferably, by an iontophoresis technique in which pilocarpine nitrate ions are transported through small areas of the skin to produce local perspiration. The chloride content may be quantitated with silver nitrate–potassium chromate-impregnated papers or with ion-selective electrodes.

Bicarbonate, phosphates, sodium, potassium, and chloride concentrations are related to maintenance of acid–base balance in the body. The pH of the blood reflects the state of the acid–base balance and is related mathematically to HCO_3^- concentration and partial pressure of CO_2 (pCO_2) in blood by the Henderson-Hasselbalch equation.

$$pH = 6.1 + \log \frac{[HCO_3^-]}{[H_2CO_3]} \tag{2}$$

Blood pH, as measured electrometrically, has a normal range of 7.36 to 7.40 for venous samples and 7.38 to 7.42 for arterial samples. The pCO_2 level in blood is determined by measuring the pH of the blood at three different pCO_2 concentrations—one native to the blood and the other two obtained by equilibration with gas

mixtures of known pCO_2. Blood bicarbonate levels also may be determined by measuring the amount of acid neutralized by plasma or serum and pCO_2 calculated by Equation 2. The relationship between pCO_2 and carbonic acid concentration is

$$\underset{\text{mM/L}}{[H_2CO_3]} = 0.03 \times \underset{\text{torr}}{pCO_2} \tag{3}$$

The role of oxygen and hemoglobin in respiration has been discussed previously. Measurements of blood pH and CO_2 content are used in differentiating respiratory acidosis (low pH, high CO_2) from metabolic acidosis (low pH, low CO_2).

Blood oxygen (pO_2) and *percent oxygen saturation* are measured by a polarographic method; the blood sample is placed in a chamber and separated from a combined platinum and silver–silver chloride electrode by a polypropylene membrane. By diffusion through the membrane, equilibrium is established between the pO_2 of the blood and a film of solution in contact with the electrode. A current, which is proportional to blood pO_2, is generated after the application of a polarizing voltage.

Calcium and *phosphorus* are important minerals in the processes of bone calcification, nerve irritability, muscle contraction, and blood coagulation. Calcium is present in plasma as an ultrafilterable (ionic and nonionic) form and a protein-bound fraction. Blood phosphorus consists of inorganic phosphorus, organic phosphate ester (G6P, ATP), and phospholipids.

Serum and urine calcium levels are determined routinely by titration with EDTA or EGTA using a fluorescent calcein or calcichrome indicator. Other methods are based on the colorimetric analysis of calcium–methylthymol blue complex in the presence of 8-quinolinol to prevent interference by magnesium. Bis-(*o*-hydroxyphenylimino)ethane forms a colored complex with calcium; in the presence of polyvinylpyrrolidone to inhibit phosphate interference, it is a sensitive and specific method for calcium. Total calcium is determined best by atomic-absorption spectrometry. As with all cations, calcium can be determined by emission- or absorption-flame photometry or ion-selective electrodes.

Inorganic phosphorus levels are determined by reaction with acid molybdate reagent to form phosphomolybdic acid which, in turn, is reduced with aminonaphtholsulfonic acid or *p*-dimethylaminophenol sulfate to give a blue complex which is estimated colorimetrically. Normal serum levels are 2.5 to 4.5 mg P/dL and 9 to 11 mg Ca/dL.

Calcium levels are decreased and phosphorus increased in hypoparathyroidism; an opposite effect is seen in hyperactivity of this gland. In rickets and osteomalacia, the concentrations of both elements are decreased. In establishing primary hyperparathyroidism and other causes of hypercalcemia, daily measurements for ionized calcium (Ca^{2+}) are replacing total Ca measurements using ISE technology.

Magnesium is an essential electrolyte that is a natural calcium channel blocker. This ion is involved with cardiac and vascular smooth muscle contraction. Hypomagnesemia may pose a risk to humans in terms of increased cardiovascular disease (such as cardiac arrhythmias and stroke), whereas hypermagnesemia may result in bradycardia, asystole, or respiratory insufficiency. A chemical analyzer (*Nova Biomedical*) is capable of measuring in 1 min Na, K, Ca, Mg, and the hematocrit in less than 200 μL of blood. Using an ion-specific electrode, the free, unbound magnesium ion activity is measured and may be related to intracellular levels of magnesium ions.

Copper, zinc, and iron are trace elements in blood. They are quantitated readily by flame photometric, colorimetric, or atomic-absorption techniques.

ORGAN FUNCTION TESTS

The analyses of various blood or urine constituents, determination of metabolic excretion rates of exogenous compounds or endogenous metabolites, and effect of exogenous stimuli on these parameters are used for evaluation of *in situ* activity and function of various organs. Organ function studies are performed in

diseases associated with the liver, kidney, parathyroid, thyroid, pituitary gland, gastrointestinal tract, pancreas, adrenals, and gonads. The principles and significance of the analysis used in such evaluations have been described also in other sections of this chapter.

Tests for *hepatic function* are based on bilirubin metabolism and excretion, carbohydrate metabolism (galactose tolerance test), plasma-protein changes (cephalin flocculation test and A/G ratio), abnormal fat metabolism, detoxification mechanisms (hippuric acid synthesis), excretion of injected substances (BSP), prothrombin formation, and previously discussed enzyme levels.

Diseases of the liver are due to cellular alterations (hepatocellular) or obstructions to the flow of bile (obstructive jaundice). Hepatocellular liver disease can be chronic (postnecrotic cirrhosis, carcinoma) or acute (viral hepatitis, alcoholism, toxin- or chemical-induced).

The *cephalin flocculation* test is based on the flocculation of cephalin-emulsified cholesterol by γ-globulin. In normal serum an albumin-like protein will inhibit this reaction; in hepatic diseases, which produce abnormal γ-globulin or reduced albumin levels, the flocculation will occur.

The *detoxification mechanisms of the liver* can be evaluated by intravenous administration of sodium benzoate and estimation of the benzoic acid metabolite, hippuric acid, in the urine. In hepatoparenchymal disease, a reduced capacity of the liver to form hippuric acid by conjugation of glycine and benzoic acid is observed.

The ability of the liver to excrete an injected dye is determined in the *BSP test*; the serum is analyzed for dye concentration at a suitable time interval after intravenous administration of 2 to 5 mg BSP/kg. Radioiodinated (^{131}I) Rose Bengal Sodium dye also has been used in dye-excretion studies with isotopic estimation of urine dye levels.

Kidney function tests are based on the determination of blood nonprotein nitrogen (urea, uric acid, and creatinine), electrolytes, blood acid–base balance, routine urinalysis, and the clearance of administered compounds in the urine. Most *clearance studies* are performed with substances that are not resorbed or secreted by the renal tubules: inulin, mannitol, sodium *p*-aminohippurate, or ^{125}I-iothalamate sodium (sodium 5-acetamido-2,4,6-triiodo-*N*-methylisophthalamate). These are administered intravenously and the rate of urine clearance and glomerular filtration is estimated by analysis of the urine. The excretory capacity of the renal tubular epithelium can be determined by measuring the clearance rate of PSP. The dye is injected intravenously and the rate of its clearance in urine is determined. PSP is bound loosely to serum albumin and is removed rapidly from the blood by the renal tubules.

Sodium iodohippurate-(^{125}I), which is extracted almost completely from the blood on a single passage through the kidney, also has been used in renal function studies; a *renogram* or isotopic scan of both kidneys is performed. The test provides data on renal tubular secretion, renal vascular competence, and renal evacuation and is primarily useful as a comparison of individual kidney function. It is important to note that 50% of kidney function can be compromised without any significant change in the routine renal function parameters.

Thyroid function tests usually measure the circulating levels of the thyroid hormones, and not the end-organ effect. The thyroid gland converts inorganic iodide to *thyroxine (T_4)* and *triiodothyronine (T_3)*. T_3 and T_4 are stored in the colloid part of the gland as part of the thyroglobulin molecule. Hypothalamic *thyrotropin-releasing hormone (TRH)* mediates the release of the pituitary thyrotropin (*thyroid-stimulating hormone, TSH).* Excess levels of circulating T_4 depress, and low levels of T_4 increase, TSH release. TSH stimulates the proteolytic degradation of thyroglobulin to release T_4 and T_3, and increases organification of iodine. T_4 accounts for 90% of secreted thyroid hormones and exists in blood bound to *thyroxine-binding globulin (TBG)* or *thyroxine-binding prealbumin (TBPA)* or to albumin. T_3 is not protein-bound and has 5 to 10 times the biological po-

tency of T_4 on a weight basis. Therefore, T_4 represents the major part of protein-bound iodine (PBI). The level of *free thyroxine (FT4),* the active fraction in blood, is regulated by T_4 and T_3 release and the levels of binding proteins in blood and tissues.

The uptake of orally administered Na ^{131}I preparations by the thyroid gland can be estimated by isotopic scanning of the gland 24 hr after ^{131}I administration and is an index of glandular function (hyperactive, $> 50\%$ uptake; hypoactive, $<15\%$).

PBI determinations are based on the precipitation of protein-bound thyroxine, removal of inorganic iodine by basic- or anion-exchange chromatography, alkaline incineration to convert thyroxine to inorganic iodide, and finally quantitation of iodide by reaction with arsenous acid and ceric ammonium sulfate. PBI is a good estimate of total circulating hormonal iodine. The normal range is 4 to 8 µg/dL serum.

T_4 can be determined by column chromatography in which it is separated and isolated by ion-exchange chromatography, and then analyzed colorimetrically. Nonisotope thyroid assays have been developed using fluorescence polarization methods for T_4 and free-thyroxin index. In the competitive protein-binding assay for T_4, serum T_4 competes with ^{125}I-T_4 for binding sites on a known amount of TBG. The ratio of bound to free ^{125}I is determined by adsorption of ^{125}I-T_4 not bound to TBG on an anion-exchange resin embedded in a polyurethane sponge or a porous dextran gel, and is a direct index of T_4 levels. The presence of mercurials, inorganic iodide, or iodinated radiographic compounds in serum interferes with the T_4 column and PBI procedures. The competitive-binding procedure is affected by the presence of highly protein-bound drugs or changes in TBG levels in serum. The normal range of serum T_4 is 2.9 to 6.4 µg/dL by column and 3.0 to 7.0 µg/dL by binding assay. T_4 and PBI are increased in hyperthyroidism and the early stages of hepatitis. T_4 and PBI are decreased in hypothyroidism and nephrosis.

FT4 also is determined in a competitive protein-binding assay in which ^{125}I-T_4 and serum are incubated, and then dialyzed to determine the percent dializable ^{125}I-T_4. FT4 analysis is used in suspected abnormalities in protein-binding globulins. T_4 binding capacity of serum TBG, albumin and prealbumin can be determined after electrophoretic separation of these proteins.

T_3 analysis is determined by the resin-uptake test. The uptake of ^{125}I-T_3 by a resin is determined in the presence of the test serum. In hyperthyroidism, the primary TBG-binding sites are saturated and ^{125}I-T_3 is taken up by the resin. The resin uptake is decreased in hypothyroidism, and most of ^{125}I-T_3 is bound to TBG in serum. A *free thyroxine index* can be obtained by multiplying T_3 (resin) $\times$ T_4 (competitive binding) $\times$ 0.01. This product deviates from normal in the same direction as T_3 and T_4 in hyper- and hypothyroidism. This product is stable during euthyroidism in spite of changes in binding proteins; for example, a euthyroid patient on phenytoin therapy will show a decreased TBG and T_4 and increased T_3, but $(T_4 \times T_3)$ is normal. The indication of hyper- or hypothyroidism in the presence of abnormal amounts of TBG is observed in the $(T_4 \times T_3)$ product.

The determination of *TSH* by RIA or EIA appears to be the most useful test in discriminating patients with primary hyperthyroidism from the euthyroidism or hypothyroidism secondary to pituitary disease. Serum TSH is increased in the primary disease state.

The *PBI conversion ratio* is an estimate of the rate of conversion of inorganic iodide to PBI. Radioiodide-(^{131}I) is administered to the subject; after 24 hr, a sample of blood is obtained and the ^{131}I to PB^{131}I is estimated by radiochromatographic procedures with ion-exchange resins (normal conversion, 13–42%).

Adrenocortical function is evaluated by estimation of serum or urinary 17-ketosteroids (17-KS) and 17-hydroxycorticosteroids (17-OH-CS) (androgen and corticosteroid metabolism), serum electrolytes (aldosterone metabolism), and blood adrenocorticotrophic hormone (ACTH) levels in the basal state, after stimulation with intramuscular or intravenous ACTH, or after adrenal inhibition with dexamethasone. In the normal individual, ACTH will increase plasma cortisol and urine 17-OH-CS, and dexamethasone will suppress plasma cortisol. Metapyrone, an inhibitor of 11β-hydroxylase, will cause selective secretion of compound S (11-deoxycortisol) by the adrenals in place of cortisol. Compound S will not inhibit the adrenal-pituitary feedback mechanism; the pituitary will secrete more ACTH and the adrenal will secrete more compound S. The determination of urinary 17-OH-CS or tetrahydro-compound S (THS) following metapyrone administration is a good index of the functional integrity of the pituitary-adrenal axis; patients with virilizing adrenal hyperplasia excrete excessive THS due to an 11β-hydroxylase defect.

Common clinical chemistry reference values are listed in Table 32-4.[18]

AUTOMATED ANALYSIS

The automation of analytical techniques used in blood and urine chemistry, hematology, blood typing, and immunology has increased the productivity and accuracy of the clinical laboratory.[19] *Computerization* of the automated analytical system also has increased the rapidity of reporting test results, reduced clerical error, and provided a unified and updated report of the laboratory tests for each patient.

In one of the first widely used multichannel chemistry analyzers, the SMA-6 (Sequential Multiple Analysis) Autoanalyzer (*Technicon*), a continuously operating, multiple-channel proportioning pump moved the samples, diluents, and reagent streams. Air bubbles segmented the flowing streams of samples and reagents, which then either flowed through dialyzers to remove interfering substances, and/or moved them directly into chambers preset at desired temperatures; they finally went into detection devices (colorimeters, fluorometers, flame photometers, spectrophotometers). A serum standard was run simultaneously with the samples. The results could be read directly from a recorder or were later coupled into a digital computer for numeric output. Sequential, multiple analyses in the SMA-12, a later model for 12-parameter analysis, was accomplished by distributing the sample to 12 different analytical streams, so that all 12 analyses were in progress at the same time. The *SMA-12 profile* usually determined calcium, inorganic phosphorus, glucose, BUN, uric acid, cholesterol, total protein, albumin, total bilirubin, alkaline phosphatase, LDH, and AST. SMA analyzers ushered in the era of automation in the clinical chemistry laboratory, spawning a high-technology industry that resulted ultimately in tremendous productivity gains, enhanced test accuracy and precision, and the ability to measure numerous constituents from microsamples in a very short time.

Typical high-throughput chemistry analyzers using standard *wet* reagent methodologies in combination with ion-specific electrodes (*ISE*) include the Synchron CX analyzers (*Beckman Coulter*), BM/Hitachi and the COBAS analyzers (*Roche Diagnostics*), and Dimension analyzers (*Dade Behring*). The non-ISE portions of these instruments basically are automated spectrometers, with robotic pipetting from on-board reagents to disposable or rewashed reaction cuvettes whose products are monitored continuously at one or more wavelengths. Stable calibration, sophisticated quality-control monitoring, off-hour preprogrammed maintenance, and self-diagnostics are regular features on these analyzers.

Vitros series analyzers (*Ortho Clinical Diagnostics*) have successfully implemented the use of "dry-slide" technology in which multilayered pads are impregnated with reagents, eliminating the need for extensive tubing and liquid reagent-handling components. Reaction end-products are measured using reflectance spectrometry in these analyzers, and analytical performance rivals or surpasses that of the more "conventional" instruments listed above.

Clinical chemistry technological innovations have enabled the laboratory to provide therapeutic drug-monitoring tests,

Table 32-4. Reference Values[a]

Electrolytes	Cortisol (free) in urine	20–90 µg/24 hr	55–248 nmol/24 hr		
Calcium	9.0–10.6 mg/dL	2.25–2.65 mmol/L			
Chloride	98–109 mmol/L	Follicle-stimulating hormone (FSH)	(adult males)	(adult females)	
CO_2 content	23–30 mmol/L				
Magnesium	1.2–2.4 mEq/L	0.6–1.2 mmol/L	2–15 mIU/mL	Follicular phase 3–15 mIU/mL	
Phosphorus	2.5–5.0 mg/dL	0.81–1.62 mmol/L			
Potassium	3.7–5.3 mmol/L	Ovulatory spike 10–50 mIU/mL			
Sodium	138–146 mmol/L				
Metabolites					Luteal phase 3–15 mIU/mL
Bilirubin	0.1–1.2 mg/dL	1.7–20.5 µmol/L			
Cholesterol	150–250 mg/dL	3.9–6.5 mmol/L			Postmenopause 30–200 mIU/mL
Creatinine	0.7–1.5 mg/dL (adults)	62–123 µmol/L	17-Hydroxycortico-steroids in urine	3–10 mg/24 hr	
Glucose	60–95 mg/dL	3.33–5.28 mmol/L			
Iron	50–165 µg/dL	9.0–29.5 µmol/L	17-Ketosteroids in urine	5–15 mg/24 hr	(adult females)
Triglycerides	20–180 mg/dL	0.22–1.98 mmol/L			
Urea nitrogen (BUN)	8–26 mg/dL	2.9–9.3 mmol/L		8–20 mg/24 hr	(adult males)
Uric acid	2.5–7.0 mg/dL	0.15–0.41 mmol/L		0.1–3.0 mg/24 hr	(prepubertal children)
Proteins and enzymes					
Alanine aminotransferase	(ALT, SGPT)	5–40 U/L at 37°	Luteinizing hormone (LH)	(adult males)	(adult females)
Albumin	3.5–5.0 g/dL	35–50 g/L		5–25 mIU/mL	Follicular phase 5–30 mIU/mL
Alkaline phosphatase	35–120 U/L at 37° (adults)	50–400 U/L at 37° (children)			
Amylase	60–180 Somogyi units	110–330 U/L			Ovulatory spike 50–150 mIU/mL
Aspartate aminotransferase	(AST, SGOT)	8–40 U/L at 37°			Luteal phase 5–40 mIU/mL
Carcinoembryonic antigen (CEA)	<2.5 ng/mL	<2.5 µg/L			Postmenopause 30–200 mIU/mL
Creatine kinase (CK)		10–180 U/L at 37°	Metanephrine in urine	<1.3 mg/24 hr	
γ Glutamyl transferase (GGT)		5–40 U/L at 37°	Prolactin	1–20 ng/mL (males)	1–25 ng/mL (females)
Lactate dehydrogenase (LDH)	60–220 U/L at 37°	(lactate → pyruvate)		(1–20 µg/L)	(1–25 µg/L)
			Thyroxine (Ty)	5.5–12.5 g/dL (adults)	7.8–16.0 µg/dL (newborns)
Total protein	6.0–8.0 g/dL	60–80 g/L			
Hormones				(72–163 nmol/L)	(101–208 nmol/L)
Cortisol in plasma	7–20 µg/dL (at 8:00 AM) (200–550 mmol/L)	3–13 µg/dL (at 4:00 PM) (80–360 mmol/L)	Vanillylmandelic acid (VMA) in urine	<6.8 mg/24 hr	

[a] Serum specimens unless otherwise indicated.
From Statland BE. *Clinical Decision Levels for Lab Tests.* Oradell, NJ: Med Econ, 1983.

drugs-of-abuse testing, and all manner of hormone analyses routinely. The advent of automated polymerase chain reaction (PCR) and other nucleic acid amplification technologies promises the development of new assays for highly sensitive detection of nucleic acid sequences, which will provide the clinician with a whole new spectrum of diagnostic information.

The hematology laboratory has kept pace with advances in automation. Multichannel analyzers sample directly from blood tubes after identifying their origin by scanning the barcode label. This "closed-sample" approach minimizes operator exposure to infectious agents that may be present in the blood. Besides performing the standard 8-parameter complete blood count, these instruments perform flow-through differential and reticulocyte counts and provide extensive graphical outputs of cell size and complexity while flagging out-of-expected-range results. Examples of these instruments include the Sysmex SE 9500 (*TOA Medical*), Coulter GEN-S (*Beckman Coulter*), AD-VIA 120 (*Bayer*), and Abbott Cell-DYN 4000 (*Abbott*). New coagulation analyzers sample plasma directly from centrifuged blood tubes and offer full menus of tests, including factor assays. An automated workstation for urinalysis (Yellow IRIS) combines standard specific gravity and semiquantitative test-strip analyses with real-time flow-through video microscopy of the urine specimen's formed elements, enabling the operator to identify and quantify cells, casts, and bacteria rapidly and precisely. In this era of health care cost containment, innovations in laboratory automation will continue to play a leading role in maximizing diagnostic effectiveness.

Urine

The formation of urine and its excretion are critical physiological activities of the body that provide a mechanism for the maintenance of a constant internal environment for all cells, tissues, and organs. This internal ecology of the body is well-recognized and known as homeostasis. Inasmuch as the urine reflects what is occurring within the body, it offers a fluid that is an important source of information, most useful as an aid in the definition of the states of health and disease. More specifically, the kidney, by means of urine formation, performs these functions:

1. Regulates the body water
2. Excretes metabolic waste products, many of which are of a nitrogenous nature
3. Excretes toxic substances of both endogenous and exogenous origin
4. Regulates the electrolyte equilibrium of the body by either excreting or retaining each specific ion
5. Maintains the delicate balance of pH within the body by excretion of excess acid or excess base
6. Provides an important route for the elimination of pharmaceutical agents and their breakdown products from the body

Normal urine contains several thousand compounds, most of which occur in minute quantities. Table 32-5 identifies some of the constituents of normal urine that are of particular significance.

Urine is studied quite widely as a means of identifying abnormalities associated with disease. The importance of such study is emphasized by the fact that the number of tests carried out on urine far exceeds those made on all other body fluids combined. Urine not only is important in providing information relating to kidney disease, but it may provide information relative to many other body activities. Information from urine studies is of diagnostic value in functional diseases of the kidney, liver, pancreas, blood, bone, muscle, and the urinary, gastrointestinal, and cardiovascular systems. Urine studies provide vital clinical information on electrolyte and water balance, acid–base equilibrium, intermediary metabolism, inborn errors of metabolism, drug abuse, intoxication, pregnancy, and hormone balance. Most of these parameters have been discussed earlier and this section will be devoted to routine urinalysis.

It is important to recognize that urine test information, like all other laboratory data, helps provide a picture of the whole body, but any single result requires interpretation to be most meaningful. It also should be recognized that negative results can be essentially as useful as positive results in a great many instances. The ready availability of urine is an advantage that makes it practical as a material for monitoring the course of the treatment of disease as well as for its recognition and definition.

Most urine examinations include observations with regard to the majority of the following: color, odor, turbidity, pH, protein, glucose (or reducing substances), ketone bodies (acetone), occult blood, bilirubin, urobilinogen, bacteria (culture or chemical tests), specific gravity, and microscopic examination of sediment, including erythrocytes, leukocytes, casts, epithelial cells, crystals, bacteria, parasites, and exfoliative cytology. A "routine" *urinalysis* varies in different institutions but ordinarily involves the inclusion of the majority of the above tests.

Urine for laboratory study should be collected in clean containers—preferably into a disposable unit (polystyrene tube) with a capacity of 15 mL that can be used for collecting, transporting, centrifuging, and testing. Refrigeration is desirable for any specimen that is not tested within 1 to 2 hr.

If urine is to be transported through the mail or is to be held for a significant time at room temperature, it is desirable to add a urine preservative (formalin, methenamine, thymol, toluene) that will interfere with microbial growth in the specimen. Several proprietary urine preservative tablets are available. If urine is allowed to stand at room temperature, bacteria will grow in the specimen and cause degradation of many constituents. Frequently, the bacteria decompose urea into ammonium carbonate with a resulting increase in the alkalinity of the specimen. Formed elements, particularly casts and red blood cells, disintegrate in alkaline solution.

The majority of urine tests are done on random specimens, but in certain instances it is necessary to have a 24-hr specimen for certain specialized analyses. For urine-sugar testing in diabetes detection, it is desirable to use a postprandial urine specimen (ie, after a meal). For protein tests, as well as chemical or culture tests for bacteriuria, the first morning specimen is preferred. Most laboratories use commercially available, standardized, reagent-impregnated strips ("dipsticks") or tablets (*Bayer*) for routine urinalysis.

INSTRUMENTATION IN URINALYSIS

Automated urine-testing systems, semi-automated reagent-strip readers, and a system that performs the complete urinalysis procedure have been developed. The strip reader is a reflectance photometer that measures urine pH, protein, glucose, ketones, blood, bilirubin, nitrate, and urobilinogen. The Yellow IRIS (International Remote Imaging Systems) measures urine specific gravity by refractometry on one channel and urine sediment on another channel by staining, passing formed elements through a flow cell, and storing individual images using video microscopy and image processing. A third channel incorporates an automated strip reader to quantify the "strip" analytes listed above. Such systems achieve more accurate and precise results for routine urinalysis than the manual or semi-automated methods.

VOLUME

The normal volume of urine excreted during a 24-hr period is usually in the range of 1000 to 1500 mL. It is possible for a healthy person to modify the volume either by severe fluid restriction or by ingestion of excessive quantities of fluid. In certain disorders there is a change in urine volume. Urine-volume increases are identified as polyuria and are encountered in diabetes mellitus, diabetes insipidus, and in certain stages of chronic renal disease. Urine volume is increased during diuretic therapy and with the ingestion or injection of large volumes of fluid. A decrease in urine volume usually occurs in dehydration, water restriction, and in acute or terminal renal disease. Extensive water loss from severe diarrhea or vomiting causes oliguria or decreased urine volume. Acute renal failure precipitated by shock, poisons, or transfusion reaction may result in a complete absence of urine excretion or anuria. In the majority of instances urine study does not require volume measurements, but these are quite critical in severely ill persons when oliguria or anuria is present.

SPECIFIC GRAVITY–OSMOLALITY

The urine density or specific gravity is related to the amount of solids excreted in a given volume of urine. In the majority of instances, in healthy persons the specific gravity varies between

Table 32-5. Normal Constituents of Urine

CONSTITUENT	G/DAY	CONSTITUENT	G/DAY
Water	1400	Amino acids	2.1
Total solids	60	Purine bases	0.01
Urea	30	Phenols	0.03
Uric acid	0.4	Proteins (total)	0.025
Hippuric acid	0.9	Chloride (as NaCl)	12
Creatinine	1.2	Sodium	5
Indican	0.01	Potassium	2
Citric acid	0.8	Calcium	0.2
Lactic acid	0.2	Magnesium	0.15
Oxalic acid	0.03	Sulfur (total)	1
Nicotinic acid	0.00025	Phosphate (as P)	1.1
Allantoin	0.04	Ammonia	0.7

1.010 and 1.030 and is related to dietary habits of fluid and food ingestion and, secondarily, to the loss of fluid by other routes such as extensive sweating. The measurement of urine density or specific gravity is a part of "routine urinalysis," and as such provides information with regard to water and solids turnover in the body. The specific gravity information alone is not nearly so important as it may be in conjunction with other observations. Thus, if dehydration is suspected, a specific gravity in the midrange of 1.015 would cast doubt about dehydration unless there was a concurrent renal dysfunction.

The kidney possesses a remarkable ability to either form a concentrated urine or a very dilute urine ranging from a specific gravity of 1.001 to 1.032. This concentrating or diluting capacity is diminished in cases of a loss of renal function. In fact, one of the sensitive tests for measuring renal function involves the so-called dilution-concentration tests where fluid is administered or withheld, and the specific gravity of the urine is measured. With a serious loss of renal function, the kidney cannot excrete a urine in excess of 1.020 even with marked fluid restriction. In advanced renal disease the specific gravity of the urine may become "fixed" or constant in the range of 1.010 to 1.012, with all urine being of this specific gravity regardless of overhydration or dehydration.

Specific gravity is measured readily with a special hydrometer, called a urinometer. There is a correlation between the density of urine and its refractive index, and a special refractometer has been designed that gives readings in specific gravity units on a single drop of urine.

Certain abnormal constituents of urine, such as glucose or protein, when present in high concentrations, will cause significant increases in specific gravity. Certain x-ray contrast media, when excreted in the urine, also will cause marked increases in specific gravity.

Urine specific gravity is only an indirect index of solute concentration; that is, 1 mol of urea will produce a lower specific gravity than 1 mol of glucose. Osmolality is a direct measure of the molal concentration of solutes in solution regardless of their molecular weight; 1 mol of NaCl dissociates into 1 mol of chloride ion and 1 mol of sodium ion. Osmolality is determined in a direct-reading osmometer by comparing the freezing point of urine with that of a standard sodium chloride solution (see also Chapter 18).

The kidneys normally excrete 800 to 1400 mOsm/kg (an osmol is that weight of any substance when dissolved in water depresses the freezing point 1.86° of solutes/day). Humans concentrate urine and eliminate the daily solute load at a maximum volume of 1200 mOsm/kg water. Urine osmolality is an inverse function of urine volume in the normal catabolic state. Urine volume is regulated by the antidiuretic hormone (ADH) and sodium excretion by the hormone aldosterone. Increased osmolality of body fluids stimulates, and increased dilution inhibits, the release of ADH. The major determinant of body-fluid osmolality is sodium. Sodium conservation is mediated through the renin-angiotensin-aldosterone axis. Determinations of plasma and urine sodium, and osmolality and urinary volume, are of diagnostic value in Addison's disease, vasomotor nephropathy (acute tubular necrosis), inapparent volume depletion, incomplete urinary tract obstruction, and hepatorenal disease.

pH

Freshly voided urine usually has a slightly acid pH. The normal range is 5 to 8, and essentially this is also the abnormal pH range. The kidneys, by reason of excreting a urine of variable pH, provide a regulatory mechanism for the body to get rid of excess acid or alkaline waste products. Because the normal pH range and the abnormal pH range are comparable, the measurement of pH alone provides minimal information, but when used in conjunction with other information, it is a very useful urinary parameter. In conditions of acidosis, the urine is quite acid; in conditions of alkalosis, the urine pH is above 7. When metabolic or respiratory acidosis is suspected, an alkaline-urine pH result almost eliminates the possibility of acidosis. Conversely, if respiratory or metabolic alkalosis is suspected, the excretion of an acid urine indicates that alkalosis is likely not present.

Dip-and-read tests are used widely for pH testing; pH-meter measurements are used less commonly. In certain situations involving kidney stone susceptibility, it is quite important to maintain a narrow range of urinary pH. For example, in cystinuria an alkaline pH is maintained to keep the cystine solubilized and to avoid as much as possible the crystallization of cystine into renal calculi. The maintenance of urinary pH is also important for optimum results in certain types of drug therapy.

COLOR

Urine normally has a yellow color, mostly due to urochrome; the color varies from pale straw to dark amber. Darker specimens usually have a high specific gravity. Occasionally, either normal or abnormal urine may show a color different from yellow. Bilirubin may cause fresh urine to be dark in color. In addition, urine that is allowed to stand darkens because of the oxidation of urobilinogen to urobilin. Red, reddish-brown, or "smoky" urine usually is due to the presence of hemoglobin (hemoglobinuria), myoglobin (myoglobinuria), or red blood cells (hematuria). Porphyria is an uncommon cause of red coloration. Black urine can be caused by melanin, which may occur in the urine of patients with far-advanced malignant melanoma. An inborn error of metabolism, alkaptonuria, is characterized by the urinary excretion of homogentisic acid, which causes the urine to turn dark brown or black on standing. Many of the unusual colors occasionally found in urine are derived from exogenous sources, including both foods and drugs. Among these are the red color caused by beets, particularly in infants, the golden-yellow or orange-red color of metabolites of pyridium-like drugs or azo drugs, and the green or blue color from methylene blue.

ODOR

Normal, freshly voided urine has a faint aromatic and characteristic odor, which is more intense in concentrated specimens. If the urine is allowed to stand, the odor becomes strongly ammoniacal and unpleasant because of bacterial destruction of urea. Freshly voided urine having a foul odor indicates severe infection. A sweet, fruity odor may be due to ketones.

APPEARANCE

Freshly voided urine is usually clear. On standing, a precipitate may form that usually consists of amorphous urates if the urine is acid or calcium and magnesium phosphates if the urine is alkaline. The formation of a precipitate is more likely to occur if the urine is refrigerated. Most specimens will become clear again if they are warmed gently to room temperature. Large quantities of mucus, cells, leukocytes, or bacteria may cause cloudiness. Protein usually does not cause cloudiness.

PROTEIN

A small amount of protein is present in the urine obtained from healthy subjects, although the quantity is not sufficient to give a positive reaction with the tests commonly used for the recognition of protein in urine. The majority of the 25 to 50 mg of protein that is excreted daily is microprotein (low-molecular-weight polypeptide), with properties quite different than those of albumin and globulin, which are the principal proteins of the blood serum. Albumin and globulins do occur in the normal urine in minute concentrations.

Plasma proteins, hemoglobin, abnormal Bence-Jones protein, and proteins (nucleo-, phospho-, and glycoproteins) derived from leukocytes and mucus may be present in urine in

nephritis, nephrosis, lesions of the urinary tract, GI dehydration, and renal congestion. Abnormal amounts of protein in the urine may be recognized by either precipitation or colorimetric tests. The precipitation depends on the heat coagulation of the protein or on the chemical precipitation of the protein. The most popular of the heat-precipitation tests is the heat- and-acetic acid test in which a tube of urine is heated to boiling after the addition of a drop or two of acetic acid. Sulfosalicylic acid is employed commonly in chemical precipitation tests; in this test, equal quantities of 3% sulfosalicylic acid and urine are mixed in a test tube, and the mixture is examined for turbidity indicative of precipitated protein.

Colorimetric tests for proteins involve *dip-and-read* type of systems and are based on the *protein error* of indicators. Certain indicators have a point of color change that is different in the presence of protein compared to the same system in the absence of protein. Thus, by buffering the indicator tetrabromophenol blue on this dip-stick at a specific pH, it is possible to have a yellow color in the absence of protein and a green or blue color in the presence of protein. This test, Albustix (*Bayer*), not only indicates the presence or absence of protein in the urine but also can be made to indicate the approximate amount of protein. Strongly alkaline or fermented urines will give false-positive results. The sensitivity of the colorimetric method is such that quantities of 10 to 20 mg of albumin/dL of urine can be recognized with confidence.

A positive test for protein in the urine may have any one of several meanings, and it is only when this information is related to other observations that it has optimum value. Proteinuria may be benign and appear following strenuous exercise or simply as a result of standing (orthostatic proteinuria). Protein frequently occurs in the urine during pregnancy and in some instances this is benign, but in other cases it indicates renal complications. Transient proteinuria may occur following severe infections, high fever, exposure to cold, and in congestive heart failure. Proteinuria may be an early and sensitive indicator of renal disease and may indicate an abnormality prior to other signs and symptoms of renal impairment in the glomerulus or tubules. In the majority of instances, there is no correlation between the amount of protein in the urine and the severity of the renal disease.

Patients with severe nephrosis may lose up to 25 g of protein/day. Such a marked loss of protein causes a decrease in plasma protein concentration with an accompanying edema. In both chronic and acute glomerulonephritis there is protein in the urine. Tumors of the kidney and renal infection usually will have an accompanying proteinuria. Bence-Jones protein is a unique protein that occurs in the urine of about 50% of patients with multiple myeloma. It has the unusual property of precipitating between 50° and 60° and dissolving at higher temperatures.

GLUCOSE (REDUCING SUBSTANCES)

Glucose normally occurs in urine in such low concentration that it escapes detection by the usual testing methods. The urine of untreated or poorly controlled diabetic patients characteristically contains easily detectable amounts of glucose. A positive test for glucose in urine usually suggests hyperglycemia and the diagnosis of diabetes mellitus; further studies, such as the glucose tolerance test to confirm the diagnosis, are indicated. Glucosuria also may occur when the renal tubules fail to reabsorb glucose normally, and glucose appears in the urine despite normal blood glucose levels, in contrast to true diabetes.

Glucose is the sugar almost always found in urine; however, lactose, galactose, levulose, sucrose, and pentoses may be encountered. These other sugars are identified by paper chromatography, selective fermentation, polarimetry, special chemical tests, or the formation of their osazones. Other reducing substances occur in urine and may cause falsely positive reducing reactions for glucose. Examples are ascorbic acid, glucuronides, many drugs, homogentisic acid, and the preservatives formalin and chloroform.

Benedict's test, the traditional test for glucose in urine, relies on the reduction of cupric ions in alkaline solution to reddish-orange insoluble cuprous oxide. The copper is reduced totally by large amounts of glucose and results in a brick-red sediment with no remaining blue color. Lesser concentrations form green to rust-colored solutions with some red sediment. A modification of this test, Clinitest (*Bayer*), is available in tablet form. The tablet contains copper sulfate, anhydrous sodium hydroxide, citric acid, and sodium carbonate. When added to dilute urine, the tablet dissolves and generates enough heat and effervescence to yield results comparable with the Benedict test.

A specific but extremely simple enzyme test for glucose is available—Tes-Tape (*Lilly*), Clinistix (*Bayer*), and Multistix (*Bayer*). Reagent strips are impregnated with glucose oxidase, peroxidase, and orthotolidine. When the stick is dipped into a solution of glucose, oxidation occurs and hydrogen peroxide is formed which oxidizes orthotolidine to a blue color. This test is more sensitive than Clinitest, but is not as reliable for estimating the concentration of glucose. The enzymatic test is specific and thus useful in determining whether a reducing substance is glucose. Diastix (*Bayer*) is a specific urine glucose test using glucose oxidase, which also indicates the quantity of glucose present.

KETONE BODIES

The ketone bodies acetone, acetoacetic acid, and beta-hydroxybutyric acid are present in the urine when fats are metabolized incompletely. Ketonuria is seen most commonly in poorly controlled diabetes and indicates ketonemia and diabetic acidosis. Other causes for ketonuria are starvation, fever, protracted vomiting, and Von Gierke's disease. Ketonuria also occurs following anesthesia. Acetoacetic acid and acetone produce a distinctive purple color when treated with a mixture of sodium nitroprusside, ammonium sulfate, and concentrated ammonium hydroxide. A similar reagent is available in tablet form (Acetest, *Bayer*). A drop of urine is placed on the tablet; if ketones are present, a lavender to deep-purple color develops in 30 sec. The color intensity indicates the concentration of ketones. The reagent strip Ketostix (*Bayer*), used as a dip-and-read test on urine or serum, contains the same reagents, which are available on Multistix (*Bayer*) and other multiple reagents as well. These tests will detect 5 to 10 mg acetoacetic acid/dL urine.

PHENYLPYRUVIC ACID

Phenylketonuria (PKU) is an inborn error of metabolism in which the normal conversion of phenylalanine to tyrosine in the body does not occur and there is a buildup of phenylalanine concentration in the blood. This metabolic disorder causes mental retardation. A portion of the phenylalanine is excreted by the kidneys into the urine and in the process is converted to phenylpyruvic acid (or phenylketone). If this genetic disorder is discovered soon after birth, it is possible to place the infant on a diet very low in phenylalanine-containing proteins and thus minimize the phenylalanine buildup in the body, averting the serious mental retardation that ordinarily is seen in the untreated PKU patient.

Recognition of PKU can be made by the use of a test for phenylpyruvic acid using a dip-and-read reagent composition containing ferric ions. This test, Phenistix (*Bayer*), can be used on urine from all newborn babies. A positive reaction gives a green color, whereas a normal infant's urine gives a pale-ivory or yellow color to the strip. PKU also can be recognized by employing a chemical or microbiological test for elevated phenylalanine in serum, as discussed under *Amino Acids*.

BILIRUBIN

Bilirubin is found in the urine of patients with hepatitis or obstructive jaundice but not in patients with hemolytic jaundice. Tests for bilirubin and urobilinogen combine to give excellent

information in the differential diagnosis of jaundice. Tests for bilirubin are of two kinds: oxidation tests form a green color of biliverdin from bilirubin usually using ferric chloride as the oxidative reagent, and diazotization tests form colored compounds when bilirubin reacts with diazonium salts in a strongly acid medium. Most oxidation tests adsorb the bilirubin onto barium sulfate or similar material before the addition of Fouchet's reagent. The tablet test Ictotest (*Bayer*) is the most sensitive diazo test, and it uses an absorption mat to concentrate the bilirubin from 5 drops of urine. A reagent tablet is added to the moist spot on the mat and 2 drops of water are added to dissolve the effervescent reagent and wash some of it off the tablet onto the mat where the reaction takes place. A blue or purple color on the mat around the tablet in 30 sec indicates the presence of bilirubin. In addition, a dip-and-read test composition also based on the diazo reaction has been incorporated into the Bili-Labstix and Multistix (*Bayer*) multiple urinalysis reagent strips. It is less sensitive than the tablet test, but its convenience allows it to be used in routine urinalysis quite readily. An incidence of approximately 0.1% positives on health-screening population groups, 0.2% on clinic patients, and 0.9% on hospitalized patients has been reported.

UROBILINOGEN

Bilirubin in the bile is reduced to urobilinogen by bacteria in the lower intestine. A portion of the urobilinogen is reabsorbed from the intestine into the blood. A portion of this urobilinogen is excreted into the urine by the kidney, and the balance is re-excreted via the bile into the intestine. Although the quantity of urobilinogen in the urine is quite small, it is an important indicator of liver function and red-blood-cell catabolism.

If there is an obstruction to bile flow such as in obstructive jaundice, the amount of urobilinogen formed and reabsorbed into the blood and excreted in the urine is decreased. With impairment of liver function, the excretion of urobilinogen in the bile is decreased, the blood concentration increases, and there is a corresponding increase in urinary urobilinogen excretion. Actually, the increase in urinary urobilinogen is one of the most sensitive tests for impaired liver function, and this test may indicate an abnormality when all other tests of liver function remain unchanged from normal.

In hemolytic diseases in which there is an increased rate of hemoglobin breakdown, the amount of bilirubin formation is increased with a corresponding increase in urobilinogen formation and excretion in the urine. The concentration of urobilinogen in urine can be established by the use of a dip-and-read test that uses the interaction of urobilinogen and *p*-dimethylaminobenzaldehyde (Urobilistix, *Bayer*).

HEMATURIA, HEMOGLOBINURIA, AND MYOGLOBINURIA

Hematuria refers to a condition in which intact red blood cells appear in the urine. This condition is indicative of a specific defect in the microscopic functional unit (the nephron) of the kidney, or it may be indicative of bleeding in the kidney, the ureter, the bladder, or the urethra. In the female there may be variable numbers of red blood cells in the urine during menstruation.

Hemoglobinuria is a condition in which free hemoglobin is present in the urine without red blood cells. This may be caused by intravascular hemolysis as a result of a transfusion reaction or by poisoning or toxins. The free hemoglobin in the plasma is excreted by the kidney into the urine. In some situations actual total hemolysis of the red cells occurs after they have entered the urine. This occurs particularly with alkaline urines.

Myoglobin is the red respiratory pigment of muscle. This pigment is quite comparable to hemoglobin in its composition and chemical reactions. Myoglobin may be liberated from muscle cells in certain types of injury and, in such cases, will circulate in the plasma and be excreted in the urine. There are also certain genetic muscle disorders in which myoglobin is lost from the muscles and appears in the plasma and subsequently in the urine.

Chemical tests for red cells, free hemoglobin, and myoglobin are based on the peroxidase-like activity of hemoglobin or myoglobin. When a chromogen mixture such as orthotolidine and peroxide is exposed to this peroxidase activity, it will interact rapidly to generate an intense blue color. A dip-and-read solid state system is available called Hemastix (*Bayer*). This specific composition uses cumene hydroperoxide as the peroxide. The same dip-and-read test for occult blood is incorporated as a component part of multiple, urine dip-and-read tests such as Multistix (*Bayer*).

MICROSCOPIC EXAMINATION

Ordinarily, urine contains a number of formed elements or solid structures of microscopic dimensions. These are studied readily by centrifuging 10 to 15 mL of urine, pouring off the supernatant, and resuspending the sediment in the drop or so of urine that remains in the tube. This suspension of sediment is placed on a microscope slide and viewed with low-power magnification. Specific structures can be studied with higher magnification. The urinary sediments can be classified into unorganized (chemical substances) and organized (cells and casts) constituents.

In an alkaline urine, amorphous or crystalline ammonium-magnesium phosphates, calcium carbonate or oxalate crystals, and ammonium urate may occur normally. Amorphous or crystalline urates, uric acid, and calcium oxalates normally are seen in acid urines. The presence of tyrosine, leucine, or cystine crystals is associated with various diseases. Chemical crystals are identified by solubility in acid and/or alkali, colorimetric reactions, and crystalline structure.

The urine sediment ordinarily contains residues of epithelial cells, crystals, and an occasional red or white blood cell. Increased numbers of erythrocytes are seen when there is bleeding into the urinary tract. If the red cells are formed into a red-cell cast, it is suggestive that bleeding has occurred at the glomerular level. An increased number of leukocytes is suggestive of infection and inflammation of the kidney. Casts are microscopic concretions that have the form of a tubule; they have a matrix of precipitated protein and, depending on their appearance, may be identified as hyaline, granular, waxy, or red-cell casts. Renal-failure casts are larger and are associated with severe necrosis of the kidney.

Numerous crystals, mucus fibers, bacteria, yeast cells, spermatozoa, and parasites (such as *Trichomonas vaginalis*) may be identified in the urine sediment. The majority of these crystals do not have any unusual significance but in certain disorders may be indicative of crystal deposits in kidney tissue or predisposition to formation of calculi.

Tissue cells can be recognized in urine sediment. This provides an excellent means of detection and diagnosis of cancer of the lower urinary tract when the sediment is fixed in alcohol and stained by the Papanicolaou procedure. Exfoliative cytology of urine may be applied as a routine to all urology patients. In one large clinic, the number of positive cases found among urology patients was almost 5%, which is a much higher return of positive results than is obtained with routine staining of cervical smears.

BACTERIA

Freshly voided specimens of urine ordinarily contain a few microorganisms, which primarily represent bacteria picked up from the external genitalia. There are fewer contaminating organisms in a *clean-catch* specimen, which involves extensive washing of the external genitalia prior to collection of the specimen. A specimen collected at the midpoint of urination or a "midstream" specimen ordinarily has more organisms than a clean-catch specimen, but fewer than a so-called random specimen.

When there is an infection of the kidney or urinary tract, the number of organisms in the urine is increased markedly. Ordinarily, if the urine contains 100,000 or more organisms/mL, the result strongly suggests the presence of an active infection. Infection of the urinary tract with accompanying bacteriuria is relatively common in young girls and women. Quite often the condition is asymptomatic and is recognized only as a result of a study of the urine. If bacteriuria is not treated, it may lead to serious renal injury.

If there is a very large number of bacteria in the urine, the specimen actually may be turbid. This can be recognized by gross visual inspection of the urine. Bacteriuria also can be recognized by microscopic examination of the urine sediment, particularly if there is a large number of organisms present. The most widely employed procedure for recognizing bacteria involves plating a specimen of diluted urine on a culture plate, incubating it, and counting the number of colonies. A more convenient approach to this same measurement involves the use of a microscope slide coated with nutrient agar. Such a slide, when dipped in a urine specimen and then incubated, will indicate the presence or absence of bacteriuria and also the approximate count.

Methods to determine the presence of significant numbers of bacteria in urine samples have been developed and incorporated into various automated systems.[20] The Bac-T-Screen (*Marion*) system was a dispensing and filtering system used with a straining process to detect the presence of bacteria on special filter cards by noting the color change on the card. Analysis on the Abbott MS-2 was performed by photometric monitoring of bacterial growth, which changed the light transmitted in a broth culture over a period of time. A decrease in the light transmission due to turbidity or color identified a positive specimen.

The Lumac Biocounter M2010 measured bacterial ATP in urine by the bioluminescence produced in a luciferin-luciferase system. Once these rapid techniques were performed to determine which specimens had increased bacteria, further identification and sensitivity testing was performed. Chemical tests for the metabolic activity of bacteria have been used in studying bacteriuria. The most popular chemical test is that for nitrite. Ordinarily, all urine specimens contain nitrate, but do not contain nitrite. If *Escherichia coli* or certain other organisms are present in sufficient numbers, they will reduce the nitrate to nitrite.

A widely used advanced automated system, the VITEK System (*bioMérieux*) has a urine identification test card that can not only detect and enumerate bacteria from urine samples, but also selectively identify the organism or organisms present using nutritionally selective components and unique metabolic indicators.

CALCULI

Knowledge of the composition of renal and bladder calculi, or "stones," is essential in planning the therapeutic regimen for such diseases. Mixed calcium phosphate and oxalate stones usually occur over the entire urine pH range. Uric acid, cystine, and calcium hydrogen phosphate calculi generally are associated with acid urines, whereas magnesium ammonium phosphate calculi usually occur in alkaline urine. Hyperexcretion of one of the calculi components, pH, renal blockage, and the presence of foreign objects in the urinary tract are the most probable causal factors in the formation of renal calculi. Calcium oxalate stones are the most common type. The chemical content of the stones is established by routine qualitative analysis for calcium, magnesium, ammonium, phosphate, carbonate, oxalate, uric acid, and cystine. Subsequent confirmation by optical crystallography, x-ray diffraction, and infrared spectroscopy is also used in the characterization of the physical properties of the calculi.

Feces

Normal feces consist of undigested food remnants, products of digestion, bacteria, and secretions of the GI tract. *Macroscopic, chemical,* and *microscopic* determinations are performed routinely. The normal quantity of feces is about 200 g/day. The brown color is a result of the reduction of bilirubin to urobilinogen and then to uribilin (stercobilin); bilirubin is not normally present in feces, but porphyrins and biliverdin (a component of meconium) are excreted during the first days of life. Bilirubin can be detected by tests previously described for bile pigments.

Color changes in the stool can be the result of dietary intake or diagnostic for biliary obstruction and gastrointestinal bleeding.[21] Patients with steatorrhea and malabsorption may show a yellow bulky stool containing fat and gas. The feces is clay colored when bile is prevented from entering the gut. A red or black stool can occur when excessive doses of anticoagulants, phenylbutazone, or salicylates are taken, producing bleeding in the gastrointestinal tract. Substances that interfere with the coloration of the stool include antacids (whitish or speckling), bismuth salts (black), iron salts (black), pyridium (orange), senna (yellow to brown), and tetracyclines (red).

Fecal urobilinogen can be determined colorimetrically by reduction of urobilin to urobilinogen with alkaline ferrous sulfate, and then reaction with acidified *p*-dimethylaminobenzaldehyde (Ehrlich's reagent). It is increased from a normal range of 40 to 280 mg/day, to 400 to 1400 mg in hemolytic jaundice (dark brown stool), and is decreased in obstructive jaundice (clay-colored stool).

Porphyrins and *porphyrinogens* do not arise from hemoglobin catabolism, such as bilirubin, but are by-products of the synthesis of heme. Increases in fecal and urinary elimination of coproporphyrin, uroporphyrin, and protoporphyrin are valuable diagnostic aids in distinguishing the various hepatic and erythropoietic porphyrias. Fecal coproporphyrins (CP) and coproporphyrinogens (CPP) are determined after extraction, conversion of CPP to CP by iodine and triple-point spectrometric estimation at 380, 401, and 430 nm to correct for interfering substances (also see section on urinalysis).

Fecal occult blood is detected readily by the *o*-tolidine, benzidine, guaiac, or diphenylamine tests; this is valid only if the patient has been on a meat-free diet for 3 days. Guaiac and diphenylamine are preferred due to the carcinogenic potential of the other two chemicals.

The Seracult test kit (*Propper*) uses an impregnated guaiac paper slide for detecting occult blood, which is a useful screening test for colon cancer. Two slides are prepared each day for 3 days from different parts of the same stool while the patient is on a meat-free high-bulk diet. Interfering substances include aspirin, indomethacin, and corticosteroids because they can produce bleeding, and vitamin C, which interferes with the oxidation reaction of the test. If bleeding occurs high in the GI tract, the blood is digested and converted to acid hematin; 50 mL of blood in the feces will cause melena (black stool). Bleeding from the lower GI tract is apparent from red streaking of stools. Also, ^{51}Cr-tagged erythrocytes have been used to quantitate and locate the source of GI bleeding. The subject's red cells are mixed with an isotonic ^{51}Cr solution and then reinjected intravenously. If bleeding occurs, the ^{51}Cr-isotope content of the feces will be increased. Location of the hemorrhagic area also can be approximated by an isotopic scan of the abdominal area.

The presence of excessive quantities of *mucus* is usually indicative of dysentery, colitis, or other inflammatory processes in the intestinal mucosa. Strongly alkaline or acidic reaction in the feces is indicative of excessive quantities of protein or carbohydrate in the diet, respectively.

Quantitative determination of *fecal nitrogen* is useful in analysis of pancreatic function. In pancreatic disease, increases in fecal nitrogen will occur as a result of decreased secretion of pancreatic proteolytic enzymes. The normal individual will excrete 4% to 13% of ingested nitrogen in the feces; in chronic pancreatitis, 9% to 30% will be excreted. Fecal nitrogen can be determined by the Kjeldahl digestion procedure.

Fecal fat is present in the form of triglycerides of fatty acids (neutral fat), free fatty acids (FFA) and soaps. Fat determinations are based on the solubility of neutral fat and FFA in ether;

the soaps are insoluble in ether and have to be acid-hydrolyzed to their respective FFA prior to extraction. Neutral fat will liberate FFA only on alkaline hydrolysis. The FFA, isolated from the above fractionations, are then determined by titrimetric, colorimetric, or gas-chromatographic procedures.

Determinations of blood, urine, and fecal ^{125}I after oral administration of an iodinated glyceryl trioleate or ^{125}I-oleic acid preparation is an index of *pancreatic, biliary,* and *intestinal absorptive function* and correlates with *fecal fat excretion*. The bile must emulsify the ^{125}I-triglyceride prior to enzymatic hydrolysis by pancreatic lipase to yield FFA-^{125}I, which subsequently is absorbed and metabolized. An increased amount of ^{125}I in the feces is associated with pancreatic diseases (cystic fibrosis with achylia), obstructive jaundice, malabsorption disease (sprue, celiac disease), and steatorrhea. The latter entity can be differentiated as to a pancreatic lipase or intestinal absorptive defect. In the "absorptive disease," increased excretion of ^{125}I is seen after administration of ^{125}I-triolein or oleic acid. In the pancreatic defect, adequate absorption of ^{125}I oleic acid occurs but fecal ^{125}I is increased after the triolein meal.

A *microscopic examination* of emulsified feces includes analysis for the presence of crystals, food residues, body cells, bacteria, and parasites. Crystals of triple phosphate, calcium oxalate, fat and cholesterol, starch granules, vegetable fibers, and neutral fat globules are normally present. Octahedral needle-shaped crystals (Charcot-Leyden crystals) are present in parasitic infestation and mucous colitis. Excessive quantities of fat or starch are seen in malabsorption disease.

Adult, larval, or ova phases of parasites may be encountered in the feces. The most common parasitic infestations are caused by *cestodes* (tapeworms), *trematodes* (flukes), *nematodes* (roundworms), and *protozoa* (amoeba) (see the section on microbiology).

Toxicology

The determination of drug or chemical concentrations in biological fluids is an important aspect in diagnosing and treating the toxic syndrome induced by various agents in acute or chronic drug-abuse situations or in chemical poisoning.

Barbiturates, glutethimide, methaqualone, chlordiazepoxide, diazepam, diphenhydramine, ethchlorvynol, morphine, phenothiazines, and salicylates are encountered in drug-abuse situations. Preliminary screening of serum or urine samples for drug substances is accomplished through the use of homogeneous immunoassay techniques EMIT (enzyme-mediated immunologic technique) or FPIA (fluorescence polarization immunoassay), or less commonly by TLC. The analysis of serum or urine levels of intact drug or its metabolites usually is performed by extraction of the sample with an organic solvent, separation by gas-liquid (GLC), or high-performance liquid (HPLC) chromatography, and quantitation by spectrometric, fluorometric, or electrochemical techniques. The technique of GC-MS (gas chromatography–mass spectrometry) methodology has become the "gold standard" because of its great sensitivity and reliability. The interpretation of the serum-concentration data in relation to clinical significance and toxicology must not be limited to numbers.

In acute drug overdosage the time of drug ingestion, time of blood or urine sampling, and severity of clinical symptoms or time of death must be interpreted in reference to data on the absorption, tissue distribution, metabolism, and elimination of the drug and its metabolites. The specificity of the chemical assay as to interference from other drugs or metabolites of the parent drug must be considered. The use of GC-MS confirms the identity of specific drugs in biological matrices. The extent of absorption of many drug substances is not related directly to the dose when large amounts of a drug are ingested, in comparison to the therapeutic dose.

The tissue-distribution and metabolic rates can be affected by large drug overdoses in which renal or hepatic failure is encountered. The plasma-elimination rate also can be affected, and it is important to recognize the change in elimination kinetics and to be aware of the nature of plasma elimination as defined by a mono-, bi-, or polyexponential elimination curve. The drug overdose usually involves several drug substances and the chemical, metabolic, and pharmacological aspects of drug interaction must be considered.

The methodology for the analysis of drugs in biological fluids or tissues can be found in the books listed in the *Bibliography*. Classic analyses for serum *barbiturate* levels will be described in this section as a specific example of the analytical methodology.

Serum is extracted at pH 6.5 with chloroform; the chloroform extract is washed with pH 7 phosphate buffer and extracted with 0.45N NaOH. The UV spectrum of the alkaline aqueous layer is determined at pH 13 and 10.5. The UV spectra are characteristic and distinguish barbiturates, *N*-methylbarbituric acids, and thiobarbiturates. The barbiturates also can be detected by acidifying the alkaline layer, extracting with chloroform, and spotting this organic extract on a silica-gel TLC plate. Sequential spraying of the plate with KMnO$_4$, HgSO$_4$, and diphenylcarbazone will show R$_f$ values and color reactions typical of the various barbiturates.

Blood barbiturates can be determined more accurately by a GLC procedure in which the retention times are used to identify the specific barbiturates. The degree of severity of clinical symptoms has been correlated with blood barbiturate levels. Comatose, areflexic signs are observed at 5.0 mg% amobarbital, 2.0 mg% pentobarbital, 8.0 mg% phenobarbital, and 1.5 mg% secobarbital.

Opiates, amphetamines, barbiturates, and methadone can be detected rapidly by "*homogenous*" *immunoassay*.[22] In this procedure, the addition of drug antibodies to a conjugate of drug and lysozyme results in the inhibition of lysozyme activity. The addition of free drug to this reaction mixture increases the enzyme activity in proportion to the amount of free drug added. The sensitivity of this type of assay is 0.1 µg/mL of amphetamine and barbiturates, 0.5 µg/mL of methadone, 0.3 µg/mL of opiates, and 1.0 µg/mL of benzoylecgonine, a cocaine metabolite. This assay is applicable to large drug-screening programs.

Electron-spin-labeling techniques also can be employed on large-scale drug-screening programs. In this procedure known amounts of drug antibodies are mixed with drug labeled with a stable nitroxide radical (spin-label) and with the specimen to be analyzed. Due to the competition for antibody between spin-labeled drug and drug in the specimen, the spin-labeled drug becomes detached from the antibody and can be detected by electron-spin resonance spectroscopy. This procedure is 1000 times more sensitive than TLC.

Blood-alcohol levels may be determined by aeration, distillation, gas chromatography, or specific enzymatic analysis with alcohol dehydrogenase. In the chemical techniques, the blood sample is either oxidized or distilled into a dichromate–sulfuric acid mixture; the excess dichromate is then determined by titration with potassium iodide or methyl orange–ferrous sulfate solutions or by colorimetric analysis. The gas chromatographic and enzyme procedures are specific for ethanol, whereas the chemical techniques are influenced by other volatile or oxidizable substances in the blood. The enzymatic method is based on the reaction of ethanol and NAD in the presence of alcohol dehydrogenase to form acetaldehyde and NADH; the acetaldehyde is removed with semicarbazide and the NADH formed in the reaction is estimated spectrophotometrically at 340 nm. Ethanol levels of >0.10% are indicative of intoxication and apparent psychomotor disturbance. Levels of 0.40% to 0.50% are associated with medullary and diencephalic disturbances such as tremors, coma, respiratory depression, peripheral collapse, and death.

Specific analysis of heavy metals is best performed by atomic absorption spectroscopy. Analyses for arsenic, beryllium, bismuth, copper, iron, lead, lithium, mercury, nickel,

thallium, and zinc are encountered frequently in the toxicology laboratory. *Blood lead* is determined by forming a lead–dithiocarbamate chelate in the presence of ammonium pyrrolidinedithiocarbamate and extracting the chelate into methyl isobutyl ketone for subsequent atomic-absorption analysis. A lead concentration of > 60 µg/mL in children usually reflects significant absorption and accumulation of lead and is interpreted as an indicator of lead toxicity (plumbism).

Increased lead exposure will result in a decrease in *delta-aminolevulinic acid* (*ALA*) conversion to porphobilinogen by ALA-dehydrase in heme synthesis. ALA blood levels will increase to the point that ALA is excreted in the urine. Determination of urinary ALA is performed by removing urine porphobilinogen and urea by ion-exchange chromatography, reacting ALA with *p*-dimethylaminobenzaldehyde and determining the chromogen colorimetrically. Urinary ALA levels > 2.5 mg/dL are unacceptable in children and industrial lead workers. Urinary ALA levels are not as sensitive an indicator of lead toxicity as blood lead, but they can be used to monitor prophylactic treatment procedures.

Cholinesterase determinations are of value in the diagnosis of suspected cases of organophosphate or carbamate pesticide poisoning. Two types of cholinesterase are found in tissues. True cholinesterase is found in RBC and nerve tissue and exhibits a specificity for acetylcholine substrate. Pseudocholinesterase is found in plasma and has a greater affinity for hydrolyzing butyrylcholine and other esters. The organophosphate and carbamate insecticides inhibit both enzymes. The activity of the plasma enzyme is inhibited more rapidly than the RBC cholinesterase, and recovers more rapidly due to synthesis of new enzyme by the liver. The recovery of the erythrocyte enzyme is slow and is governed by red-cell turnover rate. Cholinesterase activity usually is determined spectrometrically using acetylthiocholine as the substrate. Cholinesterases split this substrate into acetic acid and thiocholine which reacts with 5,5′-dithiobis(2-nitrobenzoic acid) (Ellman's reagent) to form the yellow-colored 2-nitro-5-mercaptobenzoic acid. Increasing color intensity is directly proportional to cholinesterase activity. Expected values are 3167 to 6333 U/L (serum), 1667 to 5833 U/L (plasma), and 6000 to 9167 U/L in whole blood by this methodology.

Gastric Analysis

The chief constituents of gastric juice are hydrochloric acid, gastric proteases (pepsin and gastricsin), hematopoietic factor (intrinsic factor and vitamin B_{12} binders), gastric hormones, and mucosubstances (aminopolysaccharides, mucopolyuronides, mucoids, and mucoproteins). Tests for *gastric function*[23] usually are performed on gastric juice samples collected by direct intubation into the stomach. The fasting content (normal, <100 mL) of the stomach is removed and gastric secretion is collected in the basal state, or after stimulation by the oral administration of caffeine-benzoate or alcohol, or parenteral administration of histamine, insulin, or the hormone pentagastrin. Samples are collected by continuous aspiration and analyzed for acidity and gastric protease activity at various time intervals. The extent of recovery of total juice can be estimated by oral, nonabsorbable indicators (polyethylene glycol-^{14}C, phenol red and ^{125}I-HSA) instilled into the stomach prior to the aspiration. The recovery and specific concentration of these indicators in gastric juice is an index of gastric secretory volume, completeness of collection, and gastric emptying rate.

Gastric juice is a heterogeneous mixture of clear juice and flocculent, clear mucus. The *color* of the juice should be noted as to the appearance of blood, bile, and excessive quantities of mucus. The *acidity* can be determined by a simple pH measurement and conversion to mEq of H^+ or by titration of centrifuged gastric juice to pH 3.5, 4.5 and 7.4, the respective end points for free acid (HCl), protease activity, and physiological neutrality. The *basal acid output* is about 1 mEq/hour in normal subjects

and 2 to 4 mEq/hour in duodenal ulcer patients. The *peak acid output (PAO)* after histamine stimulation is 10 to 20 mEq/hour in normals and 40 to 50 mEq/hour in duodenal ulcer; PAO following pentagastric stimulation is similar to histamine. Gastric acid secretion is decreased in atrophic gastritis, gastric carcinoma, and certain types of gastric ulcer. Hypersecretion is seen in duodenal ulcer, Zollinger-Ellison (ZE) syndrome, and hyperparathyroidism.

In situ measurements of pH may be made with a *Heidelberg capsule apparatus.* In this technique the subject swallows a small pH-sensitive capsule (transmitter); radiowaves are transmitted from the capsule to a sensing device (receiver), and the signals are recorded as a function of pH. The normal pH of the stomach is 1.2 to 1.8.

The principal gastric proteases are *pepsin* and *gastricsin;* pepsinogen is a precursor that is converted to active pepsin by free HCl and by an autocatalytic process. *Total gastric protease activity* is determined on hemoglobin or radioiodinated human serum albumin (RISA) substrates at pH 1.8 to 3.1 (RISA-^{125}I); protease activity on hemoglobin will liberate tyrosine, which can be estimated spectrometrically at 280 nm. With RISA, liberated tyrosine-^{125}I, as estimated by isotopic procedures, is an index of proteolytic activity.

Pepsin activity can be distinguished from the total protease activity by estimation of the 3,5-diiodotyrosine liberated from *N*-acetyl-l-phenylalanyl-3,5-diiodotyrosine substrate at pH 2.1.Pepsin will react on this substrate; gastricsin will not. Normal gastric juice protease activity ranges from 200 to 1200 µg total protease activity/mL and 50 to 300 µg pepsin/mL. The presence of bile, blood, saliva, or excess mucus in the sample will decrease both acidity and gastric protease activity.

Gastrin, cholecystokinin, secretin, and *pancreozymin* are gastrointestinal hormones.[24] The role of gastrin and its interaction with other gastrointestinal hormones in the etiology and proliferation of ulcer disease is of recent interest. Accurate RIA techniques have been developed for gastrin and secretin-6-tyrosine due to the availability of a pure synthetic polypeptide. Biological assays based on the effect of these substances on gastric, pancreatic, and biliary secretion also have been used.

Gastrin is found in various species in two forms, G-I and G-II. The only difference is in sulfation of the 12-tyrosyl residue in G-II of the heptadecapeptide amides. Gastrin is found primarily in the gastrin-producing cells (G-cells) of the antral mucosa. The C-terminal tetrapeptide represents the biologically active part of the molecule. Gastrin infusion will stimulate secretion of gastric acid, pepsin, and intrinsic factor. It has a slight secretin-like effect and a powerful pancreozymin-like effect on pancreatic secretion. Gastrin also stimulates bile flow. The instillation of HCl into the stomach will inhibit gastrin release; protein and meal stimulation will increase serum gastrin.

The RIA of serum gastrin is of diagnostic value in the ZE syndrome, pernicious anemia, and duodenal ulcer. Basal serum gastrin levels in the normal individual are 20 to 30 µg/mL and increase about twofold after a protein meal stimulus.

Basal serum gastrin levels in duodenal ulcer are normal or slightly elevated, but increase four- to fivefold after a protein-meal stimulus. Basal serum gastrin levels are elevated in ZE to 500 to 4000 pg/mL due to the presence of a gastrin-producing tumor. The ZE patient is uniquely sensitive to intravenous calcium stimulation, which will increase both gastric acid secretion and serum gastrin in this syndrome. Basal serum gastrin levels also are elevated in gastric hyposecretion as seen in pernicious anemia and Type A gastritis and in chronic renal failure due to the decreased metabolic turnover of gastrin in the kidney.

The RIA of serum gastrin is based on the competition of gastrin in test sample with ^{125}I-gastrin for gastrin antibody binding sites. The antibodies used in this procedure are usually cospecific for G-I and G-II. However, they detect all forms of circulating gastrin: Big-Big Gastrin (G-39), Big Gastrin (mol wt 7000; G-33), gastrin heptadecapeptide (G-17, mol wt 2200), G-13 and G-8 (mini-gastrin). The Big components can

be converted to gastrin by trypsin hydrolysis. The significance of changes in the ratio of the circulating gastrins is not known, but it has been suggested that G-39 and G-33 predominate in the basal state and cleave to G-17, which is the major serum form after a protein meal.

Other Body Fluids

Physical, chemical, and microscopic examination of cerebrospinal fluid, seminal fluid, synovial fluid, human milk, transudates, and exudates also are performed by the clinical laboratory. The principles of the various determinations are similar to those described for blood and urine.

MICROBIOLOGY

Clinical medical microbiology is a science concerned with the isolation and identification of disease-producing microorganisms: bacteria, fungi (including yeast), viruses, rickettsia, and parasites. The techniques employed in the isolation and identification of the suspect organisms involve the propagation on suitable primary culture media, selective isolation on special culture media, use of suitable living host material (mouse, embryonated egg, tissue culture, etc), determination of morphological and, where applicable, staining characteristics of the organism, and confirmation by biochemical and/or immunochemical analysis. Suitable animal inoculation, where applicable, may be employed to determine pathogenicity. Site, timing, technique (aseptic), instrumentation, and transportation of clinical specimens (blood, urine, feces, cerebrospinal fluid, etc) are prime variables involved in the final differentiation and confirmation process.

Rapid manual enzymatic and immunological test kits have been introduced to identify pathogens for cerebrospinal fluid analysis. The latex-agglutination test coats a specific antibody onto latex particles and when an antigen is present, the latex particles are visible.[25] In the coagglutination test, the specific antibody is bound to protein A on the surface of a staphylococcal cell and the presence of antigen produces agglutination.[25]

Staphylococcus aureus (*Micrococcus pyogenes* var *aureus*) is a Gram-positive coccus frequently found on normal human skin and mucous membranes and frequently associated with abscesses, septicemia, endocarditis, and osteomyelitis. Some strains elaborate an exotoxin capable of causing food poisoning. The primary isolation is on blood agar and in thioglycollate broth. With feces and other heavily contaminated specimens, phenylethyl alcohol agar and/or mannitol-salt agar should be inoculated to suppress growth of other bacteria. The identification of pathogenic staphylococci is based on colonial (pigmentation) and microscopic morphology (grape-like clusters), positive catalase production, positive coagulase production (staphylocoagulase-plasma clotting factor), and positive mannitol fermentation.

Streptococcus pyogenes is another Gram-positive coccus frequently associated with tonsillitis or pharyngitis, erysipelas, pyoderma, and endocarditis. Neopeptone agar containing 5% defibrinated sheep blood is preferred for primary isolation and to demonstrate characteristic hemolysin production by observing a zone of clear (beta) hemolysis around the colonies on blood agar. Streptococcal groups are identified by precipitin tests with group-specific antisera for A, B, C, D, F, and G. Streptex (*Diagnostic Product Corp.*) uses a latex agglutination system for identifying the Lancefield group of streptococci. Other groups usually are not associated with human clinical materials.

Legionella pneumophila identification includes specimen cultures on lung tissue or sterile body fluids (eg, pleural fluid or pericardial fluid). Direct fluorescent antibody method is a test for *L pneumophila*. Organisms are best seen in the acute stage of the disease. Because the antiserum is species-specific, polyvalent antisera are necessary for identification.

Neisseria gonorrhoeae is a Gram-negative diplococcus associated with the venereal disease gonorrhea. The identification is based on the primary isolation of the gonococcus from urethral exudates on chocolate agar or Thayer-Martin (TM) medium. The microscopic observation of Gram-negative intracellular diplococci resembling the gonococcus constitutes a presumptively positive diagnosis of gonorrhea. Confirmation of the oxidase enzyme activity of the gonococci is performed by a reaction with *p*-dimethylaminoaniline, which turns oxidase-positive colonies black. A positive oxidase test by Gram-negative diplococci isolated on TM medium constitutes a presumptively positive test for *N gonorrhoeae*. Final identification rests on typical sugar fermentation or specific (fluorescent antibody) staining.

Neisseria meningitidis is the primary cause of bacterial meningitis and septicemia. The primary isolation is based on culturing of a specimen (blood, spinal fluid, or nasopharyngeal secretions) on a Mueller-Hinton medium or chocolate agar containing a vancomycin-colistimethate-nystatin antibiotic mixture. The confirmation of the isolate by biochemical reactions (positive oxidase, positive catalase, etc) and serological agglutination with group-specific (A, B, and C) antiserum is used in the differentiation. Young cultures of groups A and C may show capsular swelling (Quellung reaction) in the presence of a specific antiserum.

The enteric bacilli (*Enterobacteriaceae*) are Gram-negative, nonsporulating rods associated with dysentery (*Shigella* spp) typhoid fever (*Salmonella typhi*), urinary tract and tissue infections (*Escherichia coli, Proteus* spp, and *Pseudomonas* spp), and pulmonary infections (*Klebsiella* spp). The primary isolation of enteric bacilli is on selective and differential infusion agar such as MacConkey and eosin-methylene blue (EMB), and enrichment media such as selenite broth and tetrathionate broth. The primary isolation of *Salmonella* spp is on Leifson's deoxycholate citrate agar (LDC) or *Salmonella-Shigella* agar (SS); if *Salmonella typhi* is suspected, brilliant green agar (BG) and bismuth sulfite agar (BS) may be used and would constitute a presumptively positive diagnosis of *S typhi*.

The confirmation and identification of enteric bacilli may be performed by serological tests and biochemical reactions: H_2S production (triple-sugar iron agar), indole production, acetyl-methylcarbinol production, citrate utilization, urease, lysine, and arginine decarboxylase and phenylalanine deaminase activity. Enterotube (*Roche Diagnostics*) employs conventional media to perform 11 standard biochemical tests that can be inoculated simultaneously in one compartmented tube, with a single bacterial colony. The serological identification of *Salmonella* and *Shigella* spp is based on the agglutination of antigens that fall into three categories: "K" capsular (*Klebsiella* spp and *Shigella* spp), "O" (*Salmonella* spp, *Arizona* spp, *E coli, Shigella* spp, etc), and "H" flagellar (*Salmonella* spp).

Other Gram-negative rods of medical importance are the hemophilic bacilli (*Bordetella pertussis*, whooping cough; and *Haemophilus influenzae,* bacterial meningitis), the hemorrhagic bacilli (*Pasteurella pestis*, bubonic plague; and *P tularensis*, tularemia), and pyrogenic bacillus (*Brucella melitensis*, undulant fever).

Spore-forming Gram-positive rods of medical importance belong to the genus *Clostridium*, which are associated with tetanus (*C tetani*), gas gangrene (*C perfringens* or *welchii*), and botulism (*C botulinum*). The isolation of these organisms requires anaerobic conditions. Once the strain to be identified is obtained in pure culture by single-colony selection, its morphological characteristics are noted; the strain then is grown in a variety of definitive media to determine catalase activity, hydrogen peroxide decomposition, and fermentation or hydrolysis of carbohydrates and organic acids. The analysis of fermentation products (gas chromatography) also is used for the identification of pathogenic anaerobic *Clostridia*. The major clostridial exotoxin type can be determined by typing with specific antitoxin sera. A Gram-positive, aerobic, spore-former of medical importance is *Bacillus anthracis*, responsible for anthrax, a disease of animals that is transmissible to humans.

The mycobacteria are acid-fast bacilli associated with tuberculosis in man (*Mycobacterium tuberculosis*) and in cattle (*Mycobacterium bovis*), and leprosy (*Mycobacterium leprae*). Tubercle bacilli in man are isolated from sputum cultured on a tubed or bottled egg medium (Lowenstein-Jensen) following enzymatic digestion and concentration of the specimens. A provisional diagnosis of tuberculosis usually is made by demonstrating acid-fast bacilli microscopically, x-ray diagnosis, and a positive tuberculin skin test.

Other weakly and partially acid-fast bacilli of medical importance are members of the Actinomycetales, *Nocardia asteroides,* and *Nocardia brasiliensis*, which are responsible for severe pulmonary infections and cutaneous and subcutaneous abscesses.

Bacteriophages (phages) are a special group of viruses that are hosted by bacteria. Any given phage is highly host-specific and when in contact, lysis of the host occurs (phage-typing). They are used primarily as epidemiological tools in subtyping strains of *E coli*, staphylococci, or *Salmonella* spp that are presumed to be related epidemiologically. Phages also furnish ideal material for studying host-parasite relationships and virus multiplication.

The medically important fungal diseases include the superficial mycoses—fungal invasion is restricted to the outermost layers of the skin or to the hair shafts (such as *Microsporum audouini*, *Trichophyton* spp, *Epidermophyton floccosum*)—and the systemic pathogenic fungi (*Blastomyces dermatitidis, Coccidioides immitis, Histoplasma capsulatum*, *Candida albicans*). The diagnosis of the causative agent is based on the isolation of organisms on Sabouraud's dextrose agar or trypticase soy agar with or without cycloheximide and chloramphenicol to suppress the growth of saprophytic fungi and bacteria, macroscopic examination of morphological characteristics, and microscopic examination using potassium hydroxide (KOH) or lactophenol cotton-blue stain. Biochemical reactions usually are limited to *Candida* spp. Immunological reactions include skin tests, where applicable; agglutination tests, such as latex particle agglutination for histoplasmosis; and tube precipitin and complement-fixation tests.

An *antimicrobial susceptibility test* is a determination of the least amount of an antimicrobial chemotherapeutic agent that will inhibit the growth of a microorganism *in vitro,* using a tube-dilution method, agar cup, or disk-diffusion method. The test may function as an aid in the selection of a chemotherapeutic agent by the physician. Also, the concentration of antimicrobial agents in body fluids may be determined by biological assay with an organism of known susceptibility for the specific agent.

The laboratory diagnosis of *viral infections* is based upon

1. Examination of the infected tissues for pathognomonic changes or for the presence of viral material
2. Isolation and identification of the viral agent
3. Demonstration of a significant increase in antibody titer to a given virus during the course of the illness
4. Detection of viral antigens in lesions, using fluorescein-labeled antibodies
5. Electron microscopic examination of vesicular fluids or tissue extracts

Blood is used for serological tests but seldom for virus isolation. Acute and convalescent-phase blood specimens must be examined in parallel to determine whether antibodies have appeared or increased in titer during the course of the disease. Some examples of human viral infections are respiratory infections (Adenovirus group); diseases of the nervous system, such as polio and Coxsackie viruses of the picornavirus group; smallpox (poxvirus group); measles (paramyxovirus group); chicken pox (herpesvirus group); and influenza (myxovirus group).

Members of *Mycoplasmataceae* pleuropneumonia-like organisms (PPLO) are of a range of size similar to the larger viruses. They are highly pleomorphic because they lack a rigid cell wall, they can reproduce in cell-free media, and they do not revert to or from bacterial parental forms as the L-forms. Spec-

imens (sputum, bronchial secretions, urinary sediment, etc) for the primary isolation of mycoplasmas (*Mycoplasma pneumoniae, M hominis,* etc) should be cultured on agar media containing peptone, serum, ascitic fluid, whole blood, or egg yolk. The species identification may be by growth inhibition on agar medium containing type-specific rabbit antisera. Antigenic variants or subspecies may be detected by immunodiffusion. Various PPLO are pathogenic, parasitic, or saprophytic. Mycoplasmas have a predilection for mucous membranes and are associated with primary atypical pneumonia and bronchitis.

Clinical parasitology is a science that is concerned with the parasitic protozoa (amoeba), the helminths (cestodes, tapeworms; trematodes, flukes; nematodes, roundworms), and the arthropods. The identification of protozoan ova is based on detailed microscopic morphology (nuclei and so on) using wet mounts (saline or iodine) or stained preparations (such as iron hematoxylin) obtained from fecal specimens (fresh or preserved with polyvinyl alcohol) that are concentrated by sedimentation, centrifugation, or flotation techniques. Trophozoite and/or cystic stages may be detected in fecal specimens associated with intestinal protozoa as in amebic dysentery caused by *Entamoeba histolytica*.

The commonly encountered helminths are *Necator americanus* (hookworm), *Trichuris trichiura* (whipworm), and *Enterobius vermicularis* (pinworm); they are identified by characteristic ova. Characterization of tapeworm segments (proglottids) or head (scolex) in a fecal specimen will differentiate *Taenia saginata* (beef tapeworm) from *Taenia solium* (pork tapeworm). Eggs of *T solium* and *T saginata* cannot be differentiated on a morphological basis.

Adult flukes oviposit a characteristic egg that may reach the urine, sputum, or feces. *Schistosoma japonicum* eggs have a small, indistinct spine; *S mansoni,* a distinct, large, lateral spine; and *S haematobium*, a distinct terminal spine.

Arthropoda is the largest of the animal phylum; arthropods are characterized by a segmented body, with the segments usually grouped in two or three distinct body regions; by a chitinous exoskeleton; several pairs of jointed appendages; and characteristic internal organs. Most arthropods can be preserved in 70% alcohol. They are of medical importance because they can infest humans and cause mechanical trauma or produce hypersensitivity from repeated exposure (eg, *Cimex lectularius*, the bedbug) or by toxin injection (eg, *Latrodectus mactans*, the black widow spider), by skin invasion (eg, *Sarcoptes scabiei*, the itch mite), and by transmitting disease (eg, *Anopheles* mosquitoes and malaria; fleas and *Yersinia pestis* or plague).

The serodiagnosis of parasitic diseases includes the following immunodiagnostic tests: complement-fixation (trichinosis), precipitin test (schistosomiasis), bentonite flocculation (ascariasis), hemagglutination (echinococcosis), latex agglutination (trichinosis), cholesterol flocculation (schistosomiasis), fluorescent antibody (malaria), and methylene blue dye test (toxoplasmosis).

IMMUNOCHEMISTRY

Clinical immunopathology[26] includes *general immunology* (immunofluorescence, immunodiffusion, immunoelectrophoresis, and agglutination tests), *radioimmunoassay* (RIA-hormones, vitamins, drugs, immunoglobulins), *tissue typing* (histocompatibility tests in organ transplants), *cellular immunology*, *cancer immunology,* and *immunohematology*. Examples of each of these disciplines are discussed in this section and other parts of this chapter.

The ELISA, *enzyme-linked immunosorbent assay*, detects antibodies by an indirect technique using enzyme-linked antibodies to label antigenic substances in tissue or body fluid. The antigen is attached to a solid matrix and reacts with a specimen that may contain a complementary antibody. The antihuman globulin, which is conjugated with the enzyme, is added and the antigen reacts with the bound antibody of the patient. By adding the substrate molecule the enzyme is detected. This analytical test system has been used to identify antibodies to

viruses, parasites, bacterial products, and in quantitation of some drugs.

Antibody response is a complex process involving the lymphoid cell system response to foreign stimulus or antigen. Hematopoietic cells in the fetal yolk sac, liver, or marrow develop into lymphoid stem cells that, in turn, differentiate into T lymphocytes of thymic origin and B lymphocytes of bone-marrow origin. The T cells further differentiate into lymphoblasts, which are responsible for *cell-mediated cellular immunity* (graft-versus-host reaction, tissue transplant rejection, tuberculin skin testing, *delayed-type hypersensitivity*). B cells differentiate into plasma cells, which are responsible for humoral immunity, which is mediated by circulating serum immunoglobulins (*immediate-type hypersensitivity*).

Macrophages can cooperate in presentation of antigen to the T or B lymphoblasts. Cooperation between T and B cells, immunological memory, development of immune tolerance to antigens, and genetic control of the immune response are integral properties of the immune system and are related to development of immune deficiency and autoimmune disease.

The identification and determination of *immunoglobulins* (IgG, IgM, IgA) by radial immunodiffusion and immunoelectrophoresis were discussed in the section on proteins. *IgM* (*γM*) is the earliest antibody found in the primary immune response and falls rapidly after the onset of IgG antibody synthesis. *IgG* (*γG*) is the major class of antibody in both the primary and secondary immune response. IgG can cross the placenta to provide the early forms of antibody protection for the newborn. IgG and IgM can participate in the complement fixation reaction. *IgA* (*γA*) is found predominantly in saliva and secretions of the gastrointestinal and respiratory tracts. In contrast to IgM and IgG, only a small portion of total IgA is found in blood. IgA functions in protection against pathogens that enter the host through the respiratory or gastrointestinal tract. *IgD* (*γD*) is found in trace quantities in sera and its function is unknown. *IgE* (*γE*) is probably the most important antibody in acute hypersensitivity or allergic reactions. Reaction of mast cell- or basophil-bound IgE with antigen initiates the release of histamine, slow-reacting substance (SRS), serotonin, and bradykinin and the subsequent allergic response. IgE is best quantitated by RIA. Mean serum levels (mg/dL) in healthy adults are IgG 1200 ± 500, IgA 210 ± 140, IgM 140 ± 70, IgD 3, and IgE < 0.1.

Heterophile antibodies are agglutinins that are capable of reacting with antigens that are entirely unrelated to those that stimulate their production. These antibodies, which occur in the serum of patients with infectious mononucleosis or serum sickness, will agglutinate formalized horse erythrocytes. To distinguish the specific *heterophile agglutinins of infectious mononucleosis*, the serum sample is mixed with guinea-pig kidney tissue or beef erythrocyte stromata; the infectious mononucleosis antibody will be absorbed and inactivated by the beef cells but not by the kidney tissue, and subsequent agglutination of horse erythrocytes will occur only in the kidney-tissue system. This test is used to detect infectious mononucleosis even prior to clinical symptoms. The heterophile titer has no relation to the course or severity of the disease.

Two protein constituents of human plasma, *rheumatoid factor* (RF) and *C-reactive protein* (CRP) are of value in the differential diagnosis of rheumatoid diseases. CRP is a protein present in the serum of patients in the acute stages of bacterial and viral infections, collagen diseases, and other inflammatory processes. The presence of this antigen in serum is detected by agglutination of polystyrene latex particles sensitized with specific CRP antibody globulin. In the management of rheumatic fever, decreases in CRP blood levels are used to measure the effectiveness of therapy.

Rheumatoid arthritis is characterized by the presence of a reactive group of macroglobulins known as RF in blood and synovial fluid. RF is a protein of the IgM globulin fraction and is regarded as an autoantibody against antigenic determinants of IgG. Analysis of RF is based on agglutination procedures employing polystyrene latex particles coated with a layer of adsorbed human gamma globulin. The RF-antibody reaction causes a visible agglutination of the inert latex particles. CRP is not elevated in rheumatoid arthritis.

β-Hemolytic streptococci, the causative agent in rheumatic fever, produce streptolysin O and S, streptokinase, hyaluronidase, desoxyribonuclease, and NADase in the body. The growth of streptococci in tissue with elaboration of these proteins serves as the antigenic stimulus to evoke the production of specific antibodies (eg, *antistreptolysin-O*, *ASO*). The quantitation of the antibody titer to these enzymes is an index of the strength of the antigenic stimulus and the extent of the streptococcal infection. These antibodies can be detected by latex agglutination (ASO) or tests dependent on the inhibition of enzyme action by the antibody (anti-hyaluronidase inhibition of hyaluronic acid depolymerization by hyaluronidase).

The laboratory diagnosis of *syphilis* (treponemal disease) and the evaluation of a chemotherapeutic approach is based on serological tests. Demonstration of an antibody-like substance *reagin,* or of true antitreponemal antibody in the serum of infected individuals is accomplished by complement fixation or flocculation tests for reagin, or immunofluorescent techniques for treponemal antibody.

In the *complement fixation* tests (Kolmer CF), reagin reacts with a complex phosphatidic acid antigen (cardiolipin) and complement; the complement is bound and will not lyse hemolysin-sensitized red cells, which were added in the second phase of the test. In normal serum the reagin-cardiolipin complex is not formed, and the complement is free to react with hemolysin and lyse the erythrocytes.

Flocculation tests for determining syphilis use a cardiolipin-lecithin-cholesterol antigen that clumps in the presence of serum reagin occurring in nontreponemal diseases and syphilis (*Venereal Disease Research Laboratory, VDRL Test; rapid plasma reagin, RPR test*).

Treponemal antibody can be detected also by the reaction of the patient's serum with treponemal antigen and subsequent confirmation with fluorescein-labeled antihuman globulin as an indicator of primary antigen-antibody reaction (*fluorescent treponemal antibody, FTA test*). The patient's serum can be treated with an extract of treponemes prior to the FTA test to remove interfering antibodies and eliminate biological false-positives (FTA-Abs test). False-positives occur in related treponematosis such as yaws, pinta, and bejel. Increased reagin titers also occur in malaria, leprosy, infectious mononucleosis, chronic rheumatoid arthritis, or systemic lupus erythematosus and in patients on hydralazine therapy.

Febrile antibodies are present in the serum of patients with certain bacterial or rickettsial infections (spotted, typhus, or Q fever). In typhus the patient's serum contains a febrile antibody that will agglutinate a suspension of *Proteus OX-19* bacteria (Weil-Felix reaction). *Salmonella* O-H, *Pasteurella tularensis,* and *Brucella abortus* antigens are used in febrile antibody tests for diagnosis of typhoid or paratyphoid fever, tularemia, and brucellosis, respectively.

Toxoplasmosis is a major cause of birth defects. An expectant mother may become infected with oocysts in uncooked meat or from cat fur and may infect the fetus transplacentally. Toxoplasmosis testing is based on detecting serum antibody by a hemagglutination procedure. Red cells sensitized by exposure to toxoplasmosis antigen are agglutinated by the specific antibody.

Radioimmunoassay (RIA)[27] has been mentioned in various sections of this chapter as an analytical tool in the measurement of hormones, immunoglobulins, drugs, and steroids. The basic principle of RIA is

$$Ag^* + Ag + Ab \rightleftharpoons Ag^*Ab + AgAb + Ag^* + Ag$$

RIA is not to be confused with the *specific reactor assay,* which uses labeled antigen and nonantibody protein receptors for vitamin B_{12}, T^4, T^3, and cortisol assays.

All procedures are based on the observation that radiolabeled antigens (Ag*) compete with nonlabeled antigen (Ag) for binding sites on specific antibody (Ab) in the formation of antigen-antibody complexes (Ag*Ab, AgAb). When increasing amounts of Ag are added to the assay, the binding sites of Ab are saturated progressively and the antibody can bind less Ag*. Therefore, the ratio of bound to free Ag* (B/F) or percent Ag* bound is a direct index of the concentration of Ag in the assay.

The requirements for RIA are (1) preparation and characterization of Ag, (2) radiolabeling of Ag, (3) preparation of specific Ab, and (4) development of the assay system and methods to separate free (Ag, Ag*) from antibody bound (AgAb, Ag*Ab) antigen.

Antigens can be prepared from natural tissue sources or preferably synthesized. ^{3}H, ^{14}C, or ^{125}I-labeled antigens are used routinely in the assay. The biological and immunochemical activity of the antigen must not be altered in the tagging procedure, and the specific activity of Ag* must be extremely high so that tracer quantities can be used in the assay. Tritium labeling and iodination (^{125}I) produce the highest specific activity, but also increase susceptibility of Ag* to internal degradation and self-radiolysis, in contrast to ^{14}C. In many instances, the original antigen cannot be iodinated, but can be altered chemically in such a way as to retain full antigenic cross-reactivity in RIA; for example, cyclic adenosine monophosphate (cAMP) has no tyrosyl or histidyl residue for iodination; ^{125}I-succinylcyclic AMP-tyrosine methyl ester retains full cross-reactivity with antibodies to cAMP and is used in the assay.

Hormones, steroids, and drug substances are *haptens*. They do not produce the antibody response when injected by themselves, but will produce antibodies specific for the hapten when injected as a hapten-protein carrier conjugate. Gastrin (hapten) is coupled to albumin (protein-carrier) by treatment with carbodiimides (CCD), which couple functional carboxyl, amino, alcohol, phosphate, or thiol groups. Morphine must be converted to the 3-*O*-carboxymethyl derivative prior to CCD coupling with albumin to provide a functional coupling group in the hapten. The hapten-conjugate usually is emulsified in a mineral oil preparation of killed *Mycobacterium* (Complete Freund's Adjuvant) and injected intradermally in rabbits or guinea pigs on several occasions. The serum antibody must have both high specificity and affinity for the antigens.

The *assay system* contains Ag*, sample-containing endogenous Ag or a standard Ag and antibody, at specified pH (6.5 to 8.5). After incubation at 5° to 37° for anywhere from 1 hour to several days, free and antibody-bound antigen must be separated. This is accomplished by *double-antibody technique, solid-phase RIA, resin techniques,* or *salt or solvent precipitation*. In the double-antibody technique, antiglobulin (Ab′) serum is added to the assay system after incubation. Ab-Ag* and Ab-Ag complexes are antibody-globulin antigen complexes. The antiglobulin will react to form insoluble Ab′-Ab-Ag* and Ab′-Ab-Ag complexes, which can be removed by centrifugation. The free Ag*, Ag is in the supernate.

The solid phase RIA is performed by coating tubes with Ab; Ag and Ag* react, compete, and bind with Ab on the wall of the tube. Unreacted Ag and Ag* are separated by decanting and rinsing the tube. Ab also can be bound covalently with isothiocyanate to dextran gel particles. Ag and Ag* will compete and bind with Ab on particles. Bound antigen then can be separated from free antigen by centrifugation.

RIA has been applied to analysis of hormones (ACTH, angiotensin I and II, gastrin, HCG, FSH, GH, glucagon, HLH, HPL, insulin, thyroxine), steroid hormones (aldosterone, androstenedione, glucocorticoids, testosterone, estrones, progesterone), drug substances (digoxin, digitoxin, amphetamines, barbiturates, morphine, LSD, ouabain), endogenous substances (cAMP, cyclic GMP, prostaglandins, immunoglobulins, hepatitis antigen, carcinoembryonic antigen—CEA). Examples of the specific assays are discussed in other sections.

CEA and *α-1-fetoprotein* (AFP) are proteins found in fetal tissue. CEA analysis was first proposed as a specific test for the early detection of bowel cancer. Although the test does not have absolute specificity for this disease, it may prove of value as a diagnostic aid and therapy monitor. CEA can be detected by RIA. Serum levels > 2.5 ng CEA/mL are found in 60% to 70% of patients with adenocarcinoma of the colon; positive levels also are found in lower percentages in carcinomas of the pancreas, stomach, liver, breast, endometrium, ovary, kidney, and bronchus, as well as in other conditions such as gastrointestinal polyps, colitis, diverticulitis, and cirrhosis. CEA appears to be associated primarily with tumors of entodermally derived epithelial tissue. The similarity between CEA and cell-surface glycoproteins and sialic acids has stimulated considerable research interest in a new approach to cancer chemotherapy.

The study of *tissue-transplantation antigens* is an important factor in studies on tissue and organ transplants. ABO blood group antigens are involved in survival of skin and renal grafts. Because of the presence of natural occurring anti-A and B, avoidance of ABO incompatibility is important in clinical grafting. The *HL-A antigens* are found on tissue and on the white cells. There is one major histocompatibility locus, comprising a number of alleles or linked genes, on a single chromosome segment. Each allele controls four to five groups of major transplantation antigens. These HL-A isoantigens affect the survival of allogenic tissue grafts and organ transplants. HL-A antigens can be typed by a leukoagglutination method in which the patient's or donor's white cells are reacted with specific HL-A antisera. HL-A typing also can be performed by a cytotoxicity test in which lymphocytes are mixed with antisera and complement. The antibody can destroy the lymphocytes if a corresponding antigen is present on the cell surface.

REFERENCES

1. Simmons A. *Hematology—A Combined Theoretical and Technical Approach.* Philadelphia: WB Saunders, 1989, p 387.
2. Christensen RL, Triplett DA. *Lab Med* 1982; 13(11):666.
3. Bollinger P, Brailas CD, Drewinko B. *Lab Med* 1983; 14:492.
4. *Central File for Rare Donors.* Milwaukee: American Association of Blood Banks, nd.
5. Lockyer WJ. *Essentials of ABO-Rh Grouping and Compatibility Testing: Theoretical Aspects and Practical Applications.* Bristol, UK: Wright, 1982, p 56.
6. Berson S, Yalow R. *Gastroenterology* 1972; 62:1061.
7. *Federal Register* 37FR17419, Aug 26, 1972.
8. Broughton PMG, Dawson JB. *Adv Clin Chem* 1972; 15:288.
9. Solberg HE, Stamm D. *Clin Chim Acta* 1991; 202(1–2):S5.
10. Meinke W. *Anal Chem* 1971; 43:28A.
11. Vidall A, et al. *Clin Chim Acta* 1991; 202 (1–2):S23.
12. Fraser CG. *Arch Path Lab Med* 1992; 116(9):916.
13. Young DS. *Effects of Drugs on Clinical Laboratory Tests*, 4th ed. Washington, DC: AAAC Press, 1995.
14. Linnet K. *Clin Chem* 1988; 34(7):1379.
15. Peterson CM. *Diagn Med* 1980; 78(Jul/Aug):73.
16. *Radial Immunodiffusion and Immunoelectrophoreses for Qualitation and Quantitation of Immunoglobulins.* DHEW Publ HSM-72-8102. Washington DC: Department of Health, Education, and Welfare, 1972.
17. Warnick GR. *Scand J Clin Lab Invest* 1990; 198:9.
18. Statland BE. *Clinical Decision Levels for Lab Tests.* Oradell, NJ: Med Econ, 1983.
19. Godolphin W, et al. *Clin Chem* 1990; 36(9): 1551.
20. Szilagyi G, Aning V, Karmen A. *J Clin Lab Automation* 1983; 3:117.
21. Bradley GM. *Diagn Med* 1980; 63(Mar/Apr).
22. Rubenstein K, et al. *Biochem Biophys Res Comm* 1972; 47:846.
23. Baron J. *Scand J Gastroenterol* 1970; 5:9.
24. Sculkes A. *Aust NZ J Surg* 1990; 60(8):575.
25. Kuhn PJ. *Mod Lab Observer* 1983; 108(Sept).
26. Sell S. *Immunology, Immunopathology, and Immunity*, 4th ed. New York: Elsevier, 1987.
27. Patrono C, Peskar BA, eds. *Radioimmunoassay in Basic and Clinical Pharmacology.* New York: Springer-Verlag, 1987.

BIBLIOGRAPHY

Alois RM. *Principles of Immunology and Immunodiagnostics.* Philadelphia: Lea & Febiger, 1988.

Balows A, ed. *Manual of Clinical Microbiology*, 5th ed. Washington, DC: Am Soc Microbiol, 1991.

Beaver PC. *Clinical Parasitology*, 9th ed. Philadelphia: Lea & Febiger, 1984.

Beck WS, ed. *Hematology,* 5th ed. Cambridge: MIT Press, 1991.

Bick RL. *Disorders of Thrombis and Hemostasis: Clinical and Laboratory Practice.* Chicago: ASCP Press, 1992.

Brostoff J, et al. *Clinical Immunology.* New York: Gower Medical, 1991.

Chandrasoma P. *Concise Pathology.* Norwalk, CT: Appleton & Lange, 1991.

Coon JS, Weinstein RS. *Diagnostic Flow Cytometry.* Baltimore: Williams & Wilkins, 1991.

Dacie J, Lewis S. *Practical Hematology*, 5th ed. London: Churchill, 1984.

Davis FA. *Modern Blood Banking and Transfusion Practices.* Philadelphia: FA Davis, 1983.

Doucet LD. *Medical Technology Review.* Philadelphia: JB Lippincott, 1981.

Edwards PR, Ewing WH. *Identification of Enterobacteriaceae*, 4th ed. New York: Elsevier, 1986.

Faulkner W, et al. *Handbook Clinical Laboratory Data.* Cleveland: Chem Rubber, 1980.

Graff L. *A Handbook of Routine Urinalysis.* Philadelphia: JB Lippincott, 1983.

Hawcroft DM. *Diagnostic Enzymology.* London: Wiley, 1987.

Henry JB. *Clinical Diagnosis and Management by Laboratory Methods*, 19th ed. Philadelphia: WB Saunders, 1996.

Hicks JM, Young DS. *Directory of Rare Analyses.* Washington, DC: AACC Press, 1997.

Kaplan A, Szabo, LL. *Clinical Chemistry: Interpretation and Techniques*, 3rd ed. Philadelphia: Lea & Febiger, 1988.

Kaplan LA, Pesce AJ. *Clinical Chemistry*, 2nd ed. St Louis: Mosby, 1989.

Lamparczyk HK. *Analysis and Characterization of Steroids.* Boca Raton, FL: CRC Press, 1992.

Lee GR. *Wintrobe's Clinical Hematology*, 9th ed. Philadelphia: Lea & Febiger, 1993.

Lynch MJ. *Medical Laboratory Technology*, 4th ed. Philadelphia: WB Saunders, 1983.

Matsuda M, et al, eds. *Fibrinogen No 4: Current Basic and Clinical Aspects* (Proc, 1989 Workshop, Tokyo), New York: Elsevier, 1990.

Melamed MR, et al. *Flow Cytometry and Sorting*, 2nd ed. New York: Wiley-Liss, 1990.

Migle JB. *Laboratory Medicine-Hematology*, 6th ed. St Louis: Mosby, 1982.

Miller LE, et al. *Manual of Laboratory Immunology*, 2nd ed. Philadelphia: Lea & Febiger, 1991.

Moffat AC. *Isolation and Identification of Drugs*, 2nd ed. London: Pharmaceutical Press, 1986.

Narins RG, ed. *Diagnostic Techniques in Renal Disease.* New York: Churchill Livingstone, 1992.

Nyhan WL. *Abnormalities in Amino Acid Metabolism in Clinical Medicine.* Norwalk, CT: Appleton-Century-Crofts, 1984.

Patrono C, Peskar BA, eds. *Radioimmunoassay in Basic and Clinical Pharmacology.* New York: Springer-Verlag, 1987.

Sonnenwirth AC. *Gradwohl's Clinical Laboratory Methods and Diagnosis,* 8th ed. St Louis: Mosby, 1980.

Stahr HM, ed. *Analytical Methods in Toxicology.* New York: Wiley, 1991.

Stockley IH. *Drug Interactions*, 2nd ed. Oxford: Blackwell, 1991.

Tiwari JL, Terasaki PI, eds. *HLA and Disease Associations.* New York: Springer-Verlag, 1985.

Walker RH, ed. *Technical Manual.* Arlington, VA: American Association of Blood Banks, 1990.

Wentworth BB, ed. *Diagnostic Procedures for Mycotic and Parasitic Infections,* 7th ed. Washington, DC: American Public Health Association, 1988.

PERTINENT REFERENCE JOURNALS

Adv Clin Chem
Am J Clin Pathol
Am Clin Prod Rev
Am J Hosp Pharm
Am J Med Technol
Anal Chem
Biotechniques
Clin Chem
Clin Chim Acta
J Clin Lab Automation
J Lab Clin Med
Lab Med
Lab Notes Med Diag
Med Lab Obs
Med Lab Tech
Scand J Clin Lab Invest
Std Methods Clin Chem

Chromatography

Mandip Singh Sachdeva, PhD

R Jayachandra Babu, PhD

The term "*chromatography*" is derived from Greek, chroma meaning, "*color,*" and graphein meaning "*to write.*" Mikhail Tswett (1906)[1] a Russian botanist used the technique to separate various plant pigments by passing solutions of them through glass columns packed with finely divided calcium carbonate. The separated species appeared as colored bands on the column and based on this phenomenon this process was called as chromatography. The chromatographic technique is now widely used for the separation, identification, and determination of the chemical components in complex mixture.

Chromatography according to USP can be defined as a procedure by which solutes are separated by a differential migration process in a system consisting of two or more phases, one of which moves continuously in a given direction and in which the individual substances exhibit different mobilities by reason of differences in adsorption, partition, solubility, vapor pressure, molecular size, or ionic charge density. The individual substances thus obtained can be identified or determined by analytical methods.[2] Thus, the term chromatography can be applied to a group of methods for separating molecular mixtures. One of the phases is a fixed bed of large surface area, whereas the other is a fluid that moves through or over the surface of the fixed phase. The components of the mixture must be of molecular dimensions, which require that they be in solution or in the vapor state. The relative affinity of the solutes for each of the phases must be reversible to ensure that mass transfer occurs during the chromatographic separation. The fixed phase is called the *stationary phase,* and the other is termed the *mobile phase.* The stationary phase may be a porous or finely divided solid or a liquid that has been coated in a thin layer on an inert supporting material. It is necessary that the stationary phase particles be as small and homogeneous as possible to provide a large surface area so that sorption and desorption of the solutes will occur frequently and efficiently. Depending on the type of chromatography employed, the mobile phase may be a pure liquid or a mixture of solutions (eg, buffers), or it may be a gas (pure or a homogeneous mixture).

Modern pharmaceutical formulations are complex mixtures including, in addition to one or more medicinally active ingredients, a number of inert materials such as diluents, disintegrants, colors, and flavors. To ensure quality and stability of the final product, the pharmaceutical scientist must be able to separate these mixtures into individual components prior to quantitative analysis. The complex nature of the polymers used in the manufacture of novel drug delivery systems makes the drug separation even more complicated. Moreover, comparison of the relative efficacy of different dosage forms of the same drug entity requires the analysis of the active ingredient in biological matrices such as blood, urine, and tissue.

Among the most powerful techniques available to the analyst for the resolution of these mixtures are a group of highly efficient methods collectively called *chromatography*. Because this technique is involved so intimately in all aspects of pharmaceutical research and development, the pharmacist or pharmaceutical scientist should possess a working knowledge of chromatographic principles and techniques. *Electrophoresis,* a separation technique especially useful for resolving mixtures of biological molecules, has some similarities to chromatography and is also discussed in this chapter.

CLASSIFICATION OF CHROMATOGRAPHIC METHODS

Chromatographic techniques can be classified into five types based on the type of equilibration process. These are (1) adsorption, (2) partition, (3) ion exchange, (4) pore penetration, and (5) affinity chromatography.

Adsorption Chromatography

The stationary phase is a solid on which the sample components are adsorbed. The mobile phase may be a liquid (*liquid-solid chromatography*) or a gas (*gas-solid chromatography*); the components distribute between the two phases through a combination of sorption and desorption processes. *Column chromatography* is a typical example of adsorption chromatography in which the solid stationary phase is packed in a tubular column, and the mobile phase is allowed to flow through the solid. *Thin-layer chromatography* is another example of sorption chromatography in which the stationary phase is a *plane*, in the form of a solid supported on an inert plate.

Partition Chromatography

The stationary phase is a liquid supported on an inert solid. Again, the mobile phase is a liquid (*liquid-liquid partition chromatography*) or a gas (*gas-liquid chromatography*). *Paper chromatography* is a type of partition chromatography in which the stationary phase is a layer of water adsorbed on a sheet of paper. In the normal mode of operations of liquid-liquid partition, a polar stationary phase (eg, water or methanol) is used with a nonpolar mobile phase (eg, hexane). This favors retention of polar compounds and elution of nonpolar compounds and is called *normal-phase chromatography*. If a nonpolar stationary phase is used along with a polar mobile phase, then nonpolar solutes are retained favoring elution of polar solutes. This is called *reversed-phase chromatography*.

Ion Exchange Chromatography

This technique uses an ion exchange resin as the stationary phase. Ion exchange resin is a polymeric matrix with the surface of which ionic functional groups, such as carboxylic acids or quaternary amines, have been chemically bonded. The mechanism of separation is based on ion exchange equilibrium. As the mobile phase passes over this surface, ionic solutes are retained by forming electrostatic chemical bonds with the functional groups. The mobile phases used in this type are always liquid. When choosing a chromatographic format for the analysis of an ionic compound, ion exchange is generally considered after attempts at developing a reversed-phase method has proven unsuccessful. However, ion-exchange chromatography is the method of choice for the analysis of inorganic ions, and it is often preferable to reversed-phase methods for the analysis of small organic ions.

Size Exclusion Chromatography

In this technique, the stationary phase is a polymeric substance containing numerous pores of molecular dimensions. The mobile phase containing analytes as solvated molecules are separated according to their size by their ability to penetrate a sieve-like structure (the stationary phase). Larger molecules that will not fit into the pores remain in the mobile phase and are not retained. This method is most suited to the separation of mixtures in which the solutes vary considerably in molecular size. The mobile phase in this type may be either liquid or gaseous. Size exclusion chromatography is used extensively for the preparative separations of macromolecules of biological origin as well as for the purification of synthetic-organic polymers.

Affinity Chromatography

This technique utilizes highly specific interactions between one kind of solute molecule and a second molecule covalently attached (*immobilized*) to the stationary phase. The immobilized molecule can be an *antibody* to a particular protein. When a crude mixture containing a large number of proteins is passed through the column, only the protein that reacts with the antibody is bound to the column. After washing all the other solutes off the column, the desired protein is dislodged from the antibody by changing the pH or ionic strength.

Capillary Electro-Chromatography (CEC)

Capillary electro-chromatography (CEC) can be defined as a liquid chromatographic method, in which the mobile phase is electro-osmotically driven through the chromatographic bed. The mobile phase in CEC has proven to be superior over other chromatographic methods in terms of its efficiency in separating ionic compounds and biomolecules.

The classifications given above for the various types of chromatographic processes can be deceptive in their simplicity. Except in isolated cases, pure adsorption or partition chromatography rarely occurs. In practice, separations frequently result from combination of adsorption and partitioning effects. The ultimate success of a chromatographic separation depends on the ability of analysts to recognize the limitations of the methods and adjust their experiments accordingly. The individual types of chromatographic techniques mentioned above are shown in Figure 33-1. The types of chromatography useful in qualitative and quantitative analysis that are employed in the USP assays and tests are Column, Gas, Paper, Thin-Layer, and High-Pressure or High-Performance Liquid Chromatography (HPLC). Paper and thin-layer chromatography are ordinarily more useful for purposes of identification because of their convenience and simplicity. Column chromatography offers a wide choice of stationary phases and is useful for the separation of individual compounds,
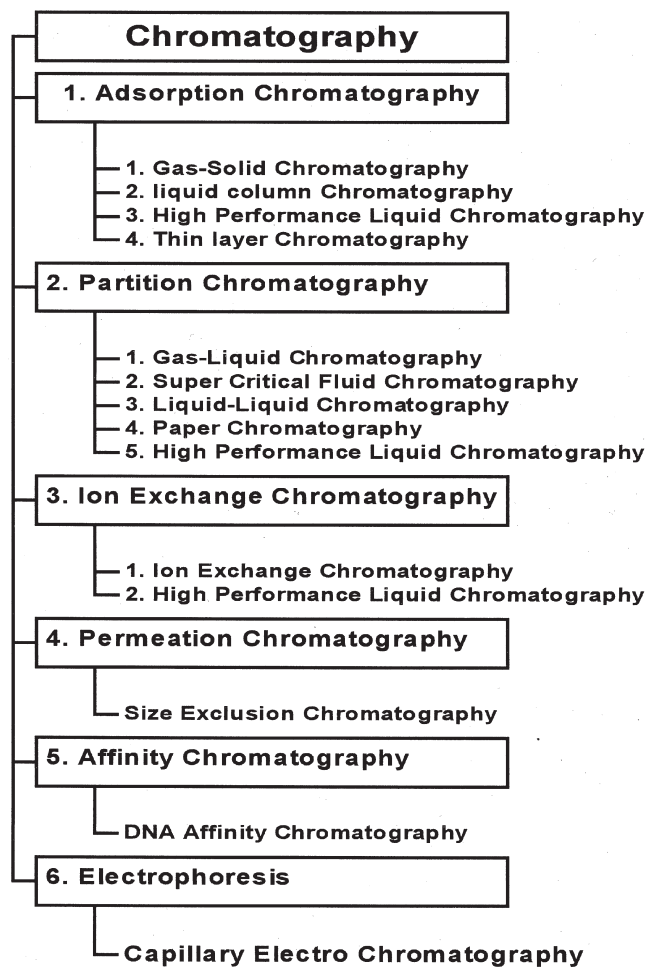
Figure 33-1. Classification of chromatographic techniques.

in quantity, from mixtures. Both GC and HPLC require elaborate apparatus and usually provide sophisticated methods to identify and quantify very small amounts of the material. A distinction needs to be made between *analytical* and *preparative-scale chromatography*. Analytical processes are used to identify and quantify tiny amounts of unknown materials. Preparative-scale chromatographic systems generally consist of a large cylindrical column within which the stationary material is packed. The mobile phase is invariably a liquid, and the stationary phase is either a solid, or a liquid supported by an adsorbent solid. Since the column is packed with stationary phase, liquid mobile phase must be forced through the column at a steady *pressure* for achieving the separation of the solutes of interest.

THE CHROMATOGRAPHIC PROCESS AND TECHNIQUES OF COLUMN DEVELOPMENT

To appreciate the theory and applications of chromatography, it is worthwhile to consider the events taking place in an ideal chromatograph. Conceptually, chromatography may be considered as being similar to the processes occurring in fractional distillation or sequential solvent extraction. In distillation, mixtures of liquids are separated by a series of steps involving vaporization and subsequent condensation. Each step involves an equilibrium between a vapor enriched in the more-volatile component and a liquid condensate of the same composition. Each single equilibration between the phases is termed a

theoretical plate, and the length of the column required for one equilibration is called the *height equivalent to a theoretical plate* (*HETP*). The nomenclature has been adopted by chromatographers to describe the equivalent transfer of solute between the mobile and stationary phases.
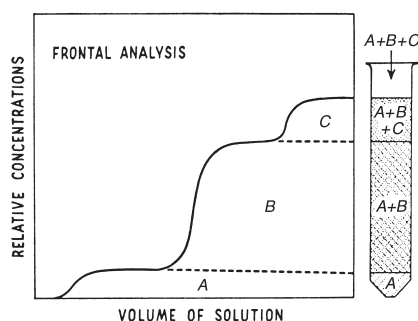
In solvent extraction, a solute, commonly dissolved in an aqueous vehicle, is transferred partially in one step into an immiscible solvent. The amount of solute transferred is determined by its partition coefficient, which is the ratio of its concentration (in reality, activity) in the nonaqueous and aqueous phases, respectively. After the first step, the layers are separated, fresh solvent is brought in contact with the aqueous phase, and as a result, a new equilibrium based on the partition coefficient is established and more solute is transferred to the nonaqueous phase. Each of these extraction steps is equivalent to one theoretical plate and is analogous to the solute-transfer process occurring in a chromatographic system.

Chromatographic processes are classified according to the physical states of the mobile and stationary phases, that is, whether they are gaseous, liquid, or solid. Each of these techniques may be classified further depending on the method of mobile-phase development into *frontal analysis, displacement analysis,* and *elution analysis.*

FRONTAL ANALYSIS

In frontal analysis a large volume of a sample mixture is allowed to flow continuously through a chromatographic column. The most weakly retained component of the mixture emerges alone from the column first (Fig 33-2). After a period of time, during which the first component elutes continually at a constant rate, a sharp front appears indicating the appearance of the next most weakly retained compound. This now elutes as a mixture with the first component. The appearance of the next front indicates the emergence of the third most weakly retained compound in a mixture with the first two. This process continues until the effluent has the same composition as the sample being introduced into the column. After this point, no further separation can occur.

Because only the component that elutes first can be obtained in a pure state, frontal analysis never has been used extensively. However, research has indicated that it may be useful for the analysis of complex mixtures that cannot be resolved by other means. If the first derivative of the frontal chromatogram is taken, the resulting graph resembles exactly a normal elution pattern. The point of maximum height of the peak for each component corresponds to the inflection point of each rising front. The flat portions of the frontal chromatogram, being constant, give derivatives of zero and thus form the baseline. The computations of the derivatives can be done easily by a computer.

Figure 33-3. Displacement development for determining the number, nature and concentration of solutes. A sample containing Solutes *A* + *B* + *C* is applied to the top of an adsorption column. The chromatogram is developed with a solvent containing a displacing agent (*D*) that is adsorbed more strongly on the column than either *A*, *B*, or *C*.

DISPLACEMENT ANALYSIS

In displacement analysis, the sample mixture, dissolved in a small volume of solvent, is introduced onto the column as a narrow band at the top. The mobile phase, containing a *displacing agent,* then is allowed to pass through the column. The displacing agent is a substance that is retained more strongly by the stationary phase than any of the components of the sample mixture, and it, therefore, forces them off the surface of the stationary phase into the mobile phase.

As each of the displaced solutes move through the column in the mobile phase, it in turn acts as a displacing agent for less strongly retained compounds. The final result is that the solute that is bound least firmly is eluted first, followed in order by those more tightly bound and finally by the displacing agent. A displacement chromatogram is illustrated in Figure 33-3. The pattern is similar to that obtained with frontal analysis except that the trailing edge of each solute zone does not extend back through the length of the column.

Although displacement analysis is not used in quantitative studies, it has two potential advantages: it is possible to isolate in a pure state at least a portion of each of the compounds eluting from the column and in the course of the separation process the sample is concentrated, instead of being diluted as usually occurs in chromatographic analyses.

ELUTION ANALYSIS

Elution analysis is carried out by introducing the sample in as small a volume as possible onto the head of the column. The mobile phase then is allowed to flow through the system. The components with larger partition coefficients will be retarded in their passage through the system and will "elute" later. A typical elution chromatogram is shown in Figure 33-4.

The advantages of elution chromatography are that each component of a separated mixture can be isolated in a relatively pure state contaminated only by mobile phase and that the method can be used readily for quantitative analysis. If the composition of the mobile phase is not changed during the course of the development of the chromatogram, the technique is called *isocratic-elution analysis.*

A widely used modification of elution analysis, which is capable of overcoming the difficulties of long elution times and poor resolution of complex mixtures, is called *gradient-elution analysis.* In this adaptation two eluting solvents, one *weak* and one *strong,* are used to develop the chromatogram. The *weak* solvent has a lower affinity for the solutes, whereas the *strong* solvent has a higher affinity. The elution begins using only the weak solvent and, as the development progresses, the concen-

Figure 33-2. Frontal analysis for determining the number of components in a mixture. A solution containing a mixture of the Solutes *A*, *B*, and *C* is percolated through the adsorption column at the right. *A* is adsorbed least strongly and appears first in the effluent solution. This is followed by a mixture of *A* + *B* and finally *A* + *B* + *C*. The elution diagram illustrates the increasing concentration of solutes in the effluent.

Figure 33-4. Elution development for separating components of a mixture. A sample containing Solutes *A + B + C* is applied to the top of an absorption column, and the chromatogram developed by percolating pure solvent (*S*) through the column. The components separate as they pass down the column and are collected separately in the effluent.

tration of the strong solvent is increased gradually until the final mobile phase has a composition approaching that of the strong solvent. The mixing of the two solvents is done in a specially designed chamber at the top of the column. The result is that the composition and strength of the mobile phase change constantly during the analysis. Weakly retained solutes are eluted first by the weak solvent, and strongly retained solutes, which would not elute at all with the weak solvent or which would have undesirably long retention times, are eluted by the increasingly stronger mobile phase.

THEORY OF CHROMATOGRAPHY

Two theoretical approaches have been developed to describe the processes involved in the passage of solutes through a chromatographic system.

The *plate theory,* based on the work of Martin and Synge,[3] considers the chromatographic system as a series of discrete layers of theoretical plates. At each of these, equilibration of the solute between the mobile and stationary phases occurs. The movement of the solute is considered as a series of stepwise transfers from plate to plate.

The *rate theory,* discussed in the book by Giddings (see *Bibliography*), considers the dynamics of the solute particle as it passes through the void spaces between the stationary phase particles in the system as well as its kinetics as it is transferred to and from the stationary phase.

Aspects of both of these theories will be presented in the following discussion in order to exemplify the basic principles underlying the chromatographic process and introduce the experimental parameters necessary for the understanding and interpretation of chromatograms.

Chromatographic systems achieve their ability to separate mixtures of chemicals by selectively retarding the passage of some compounds through the stationary phase while permitting others to move more freely. Therefore, the chromatogram may be evaluated qualitatively, by determining the R_f, or *retardation factor*, for each of the eluted substances. The R_f is a measure of the fraction of its total elution time that any compound spends in the mobile phase. Because the solute particle proceeds down the column only when it is in the mobile phase, the R_f is related directly to the fraction of the total amount of solute that is in the mobile phase, and it can be expressed as

$$R_f = \frac{V_M C_M}{V_M C_M + V_S C_S} \qquad (1)$$

where V_M is the volume of the mobile phase, and V_S is the effective volume of the stationary phase—the volume available for interaction with the solutes. The variables C_M and C_S indi-

cate the concentrations of the solute in the respective phases at any time. By dividing each term of the fraction by C_M, this can be simplified to

$$R_f = \frac{V_M}{V_M + K V_S} \qquad (2)$$

where K, the partition coefficient, equals C_S/C_M, the ratio of the solute concentration in the stationary phase to that in the mobile phase and is an equilibrium constant that indicates the differential affinity of the solute for the two phases. It can be seen from this expression that a component with a large partition coefficient—one that is attracted strongly to the stationary phase—will have a small R_f and a long elution time because only a small fraction of its total mass will be in the mobile phase at any time. By dividing each term of the fraction by V_M, an alternate expression results

$$R_f = \frac{1}{1 + k'} \qquad (3)$$

where the *capacity factor, $k' = K V_S/V_M$.* The capacity factor, which normally is constant for small samples, is a parameter that expresses the ability of a particular solute to interact with a chromatographic system. Because the volumes of the stationary and mobile phases are constant for any chromatographic experiment, k' is directly proportional to the partition coefficient. Therefore, the larger the value of k', the more the sample is retarded.

Both the retardation factor and the capacity factor may be used for qualitative identification of a solute or for developing strategies for improving separations. In terms of parameters easily obtainable from the chromatogram, the R_f is defined as the ratio of the distance from the origin traveled by the solute band to the distance traveled by the mobile phase in a particular time.

The R_f is used most conveniently in *complete chromatography,* such as paper and thin-layer chromatography, which occurs when the mobile phase is allowed to develop to a predetermined point in the system and then is stopped. Solutes then will have moved only a fraction of the distance traveled by the mobile phase. In *continuous chromatography,* as exemplified by the gas- and liquid-column techniques, mobile-phase development is permitted to continue indefinitely until the solutes elute from the end of the stationary phase. Measurement of the capacity factor, described below, is more useful in the latter cases.

The time that elapses from the start of the chromatogram to the elution maximum of the solute is called the *retention time, t_R,* a function of the length of the column and the rate of travel of the solute. The rate of travel is determined by

$$Rate = \mu\, R_f \qquad (4)$$

where μ is the linear velocity of the mobile phase, usually expressed in cm/sec. Thus,

$$t_R = \frac{Length}{Rate} = \frac{L}{\mu}(1 + k') = t_0(1 + k') \qquad (5)$$

where t_0 is the time for the elution of a solute that is not retained by the chromatographic system. From this, a convenient expression for the experimental determination of the capacity factor can be formulated as

$$k' = \frac{(t_R - t_0)}{t_0} \qquad (6)$$

The values of k' ideally should be between 1 and 10; that is, solutes should be retained from 2 to 11 times as long as the unretained compound. Values of k' greater than 10 result in longer retention times and broad peaks, while values less than 1 lead to poor separation.

Another parameter used to describe the retardation of a solute is the *retention volume, V_R,* which is equal to the volume of mobile phase required to elute a compound from the system.

Therefore, the retention volume is equal to the product of the retention time and the flow rate of the mobile phase, $t_R F$, or $t_0(1 + k')F$. Because $t_0 F$ is equal to the volume of mobile phase in the system (V_M, or the *void volume*), the retention volume can be expressed as

$$V_R = V_M(1 + k') = V_M + K V_S \qquad (7)$$

Therefore, the retention volume of a solute depends on the relative volumes of the two phases and the partition coefficient. Because the phase volumes are identical for each solute in a mixture, the most important influence on retention arises from the partition coefficient. A large partition coefficient results in long retention since the solute spends more time in the stationary phase.

The retention time and retention volume frequently vary slightly from run to run due to small changes in operating parameters such as temperature and flow rate. To minimize the errors caused by these variations, retention time and/or volume frequently are measured with respect to another peak in the chromatogram, rather than from the origin. Because the peak of interest and the reference peak are affected similarly by the changes in experimental conditions, the retention measurements are more accurate. In these cases, the parameters are termed *relative retention time, RRT,* and *relative retention volume, RRV.*

The elution pattern of an ideal chromatographic peak is a curve whose shape is Gaussian. Thus, it can be described by parameters derived from the normal statistical distribution, ie, the standard deviation, σ, and the variance, σ^2. It can be seen clearly by reference to Figure 33-5 that the peak width at any point can be expressed as a multiple of the standard deviation. The inflection points are located at one standard deviation on either side of the mean at a level that is 60.7% of the overall height of the peak. The width at this point is therefore 2σ. If tangents to the peak are drawn through the inflection points and extended to the baseline, the width at the base, W_B, is 4σ. The width at one-half the height is 2.354σ (W_H).

Two further characteristics of the peak are the height and area. The area is equal to the integral of the equation representing the curve from the point where it leaves the baseline to the point where it returns and is proportional to the amount or concentration of the solute. The height is measured at the maximum and, therefore, corresponds to the greatest concentration in the zone. It is at the point of maximum height that retention times and volumes are measured.

Two parameters commonly used for estimating the effectiveness of a chromatographic system are N, the number of theoretical plates, and H, the height equivalent to a theoretical plate (HETP), which is defined as L/N, where L is the length of the column. Because the width and standard deviation of a peak can vary depending on experimental conditions, a better indicator of the sharpness of a peak is its *relative standard deviation (RSD)*, σ/t_R. In practice, N is defined in terms of the re-

ciprocal of the RSD by the expression $N = (t_R/\sigma)^2$. Because it would be difficult to determine σ for each peak, the relationships given above ($W_B = 4\sigma$, $W_H = 2.354\sigma$) can be substituted to arrive at the equations
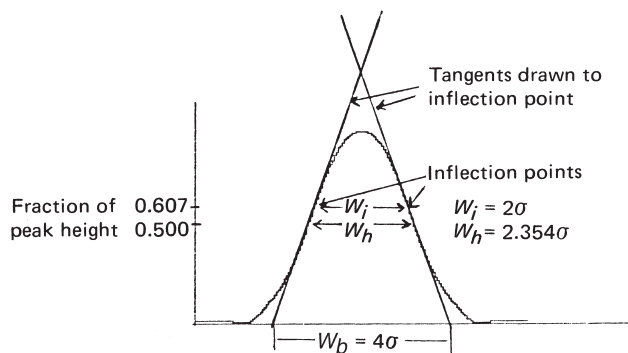
$$N = 16(t_R/W_B)^2 \text{ and } N = 5.545(t_R/W_H)^2 \qquad (8)$$

which are evaluated readily from the chromatogram. Although these are mathematically equivalent expressions, the former is used more frequently. However, the latter is useful particularly for nonideal peaks of unsymmetrical shape and possibly skewed or tailed, as the asymmetry is less pronounced at the half-height. At any particular retention time a system with a greater number of theoretical plates per unit length will produce a narrower peak and, therefore, will be capable of separating more complex mixtures.

Chromatographic systems are available in which N is 50,000/m or better. These values are established with selected test compounds, and the analyst should be aware that such levels would not be obtained with every sample. Because of intrinsic differences in the affinities of different compounds for the stationary phase, every solute will have a unique value for N in a particular system.

These procedures enable the chromatographer to derive from the experimental data a number of parameters, which characterize the retention behavior of individual compounds in a system. However, the greatest utility of chromatography lies in its ability to separate mixtures of solutes so that a number of individual substances may be quantitated or isolated in a pure state.

To develop strategies for accomplishing these objectives, consideration must be given to parameters that describe the interrelationships of both the retention- and peak-shape variables for more than one peak. The most significant of these parameters are *separation,* which is concerned with the relative positions of the band centers, and *resolution,* which describes the overlap of the leading and trailing edges of successive peaks. These are illustrated in Figure 33-6. In Figure 33-6A, a chromatogram with poor separation and resolution indicates the presence of two peaks, but is useful neither for quantitation nor for isolation of either substance. In Figure 33-6B, adequate separation has been achieved, but resolution remains poor because of overlap of the trailing edge of Peak 1 and the leading edge of Peak 2. In Figure 33-6C, the separation has remained constant while resolution has been optimized to lessen band overlap, resulting in an ideal chromatogram.

To achieve adequate separation of two adjacent peaks it is necessary to adjust the experimental variables so that the band centers or peak maxima elute at significantly different points on the chromatogram. This requires that the partition coefficients of the two solutes be sufficiently different so that one substance is retained more strongly than the other. Therefore, α, the *separation factor* or *selectivity factor,* may be defined as K_2/K_1, which is the ratio of the partition coefficient of the solute producing the second band to that of the solute producing the first. Because k', the capacity factor, is directly proportional to K, the separation factor also may be stated as the ratio of the respective k' values, k_2'/k_1'. From an experimental viewpoint, this is more useful, because k' can be determined more easily from peak retention parameters than K. Therefore, the separation factor usually is stated in terms of the *adjusted retention times* or *volumes* as

$$\alpha = \frac{(t_r)_2 - t_0}{(t_r) - t_0} = \frac{(V_R)_2 - V_M}{(V_R)_1 - V_M} \qquad (9)$$

Because it is based on RRT and RRV, the separation factor also is termed the *relative retention.* In addition to being useful in optimizing chromatographic separation, α also has value in qualitative analysis by chromatography. If, under identical experimental conditions, an unknown compound has the same relative retention as a known substance, the identity of the

Figure 33-5. Distribution characteristics of a typical Gaussian peak.
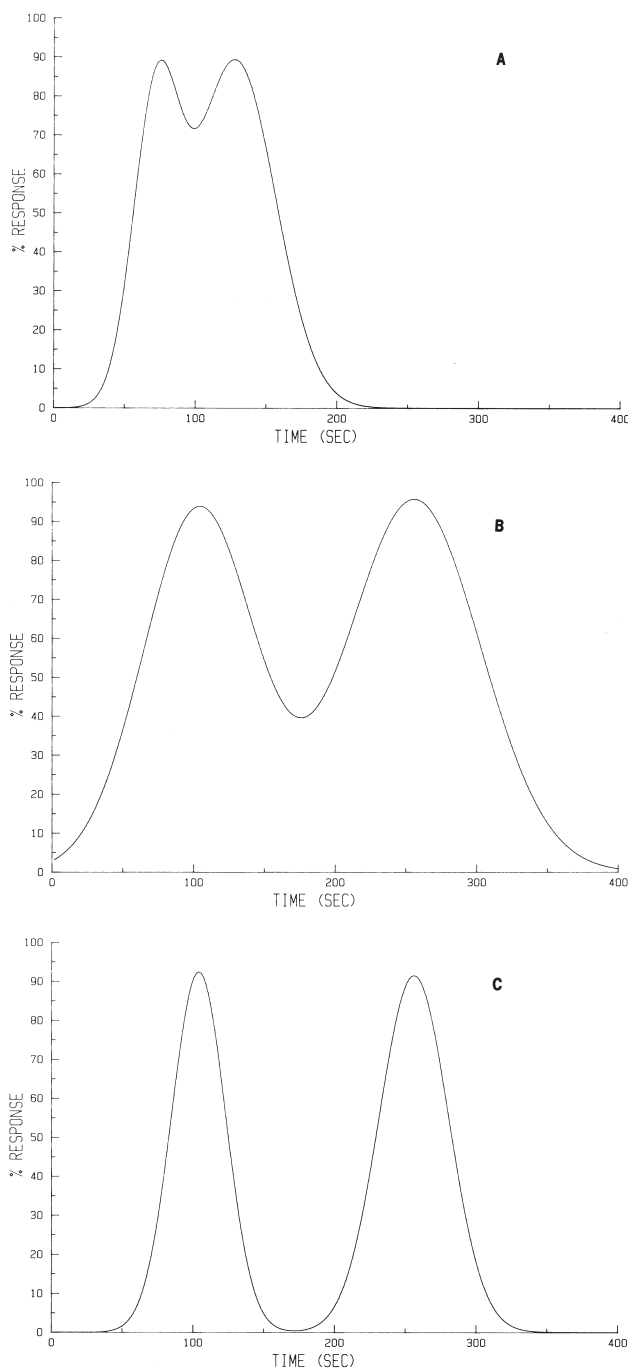
Figure 33-6. Effect of changes in separation and resolution on the elution pattern of adjacent peaks. **A.** The band centers are separated poorly and resolution is poor. **B.** Separation has been increased but band overlap remains, causing poor resolution. **C.** Peak separation is the same as in B, but overlap is reduced, giving good resolution.

number of theoretical plates is low. For the purpose of comparison the *resolution* between two adjacent peaks can be defined as the distance between the band centers divided by the average peak width

$$R_S = \frac{2(t_{R2} - t_{R1})}{W_1 + W_2} \tag{10}$$

where the peak widths at the baseline are measured by drawing tangents through the inflection points and are, therefore, taken as four times the standard deviation. For two adjacent peaks of equal size, when $R_S = 1.00$ there will be a 4% contamination of each component by the other due to overlap. At $R_S = 1.25$ the overlap will be 2%, and at $R_S = 1.50$ it will be 0.3%. The calculation of resolution is useful especially when chromatography is being used to isolate pure compounds, as it gives the chromatographer an indication of where to start and end the collection of the peak to achieve the desired purity. An alternate equation that treats resolution in terms of easily measurable experimental parameters is

$$R_S = \left(\frac{N}{4}\right)\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k'}{k' + 1}\right) \tag{11}$$

where N is the number of theoretical plates, α is the selectivity or separation factor, and k' is the capacity factor. Using this equation, the chromatographer can devise strategies for improving resolution by altering experimental conditions so as to affect one or all of N, α, or k' favorably. A more-detailed discussion of resolution, especially in those cases where the adjacent peaks are not equal in size and shape, can be found in the text by Snyder, Kirkland, and Glajch (see *Bibliography*).

TECHNIQUES OF CHROMATOGRAPHY

The five basic modes of chromatography—adsorption, partition, ion-exchange, size-exclusion, and affinity—can be applied to the analysis of pharmaceutical systems by a number of techniques that differ from each other according to the nature of the stationary and mobile phases and the apparatus used. Although it may be possible to analyze a sample using more than one of these methods, the choice of a particular technique depends on a number of factors, including the complexity of the sample, the chemical and physical properties of the compounds to be separated, the resolution required, the ease and speed of the technique and its ability to be automated, the availability and cost of the equipment, and the need to isolate the separated analytes.

If the materials are volatile and stable in the gas phase, GC may be the technique of choice because it is simple to perform, rapid, and capable of high resolution. If it is necessary to isolate eluted compounds in quantity, liquid- partition or thin-layer chromatography may be a more advantageous choice. Gas-chromatographic columns cannot handle large quantities of material and it is difficult to retrieve the eluants from the hot effluent gases. If the substances have a high molecular weight, such as proteins, triglycerides or polymers, liquid chromatography using the size-exclusion mode is necessary to achieve separation.

For compounds that are ionized in solution, such as amino acids, the ion-exchange mode of liquid chromatography particularly is useful. Highly polar or hydrophilic compounds of intermediate molecular weight, such as sugars, can be separated by partition techniques involving paper or column chromatography. Substances that are nonionizable, hydrophobic, or nonpolar are amenable to separation by liquid-adsorption methods. Highly selective isolation of certain biological substances, such as antibiotics or enzymes, may be effected using affinity chromatography. Capillary electro-chromatography can be used to analyze wide range of compounds (including chiral analytes) in biological fluids (eg, plasma, urine) in significantly shorter times than achieved by current GC and HPLC methods.

unknown substance can be inferred. However, more positive identification, with greater confidence, requires that the same relative retention for the test sample be exhibited in two *different* chromatographic systems.

As shown in Figure 33-6B, it is possible to attain adequate separation of the peak maxima and yet fail to have a useful chromatogram because of overlap of the adjacent portions of the two peaks. In this case the partition coefficients of the two compounds are sufficiently different to effect separation, but the efficiency of the chromatographic system in terms of the

GAS CHROMATOGRAPHY

In 1941, in their paper on partition chromatography, which outlined the plate theory, Martin and Synge proposed the technique of gas chromatography (GC) with the following statement:

> "The mobile phase need not be a liquid but may be a vapor. Very refined separations of volatile substances should therefore be possible in a column in which a permanent gas is made to flow over gel impregnated with a non-volatile solvent in which the substances to be separated approximately obey Raoult's Law."[3]

In the subsequent 10 years, no one followed up on this suggestion, so Martin returned to it himself and with James developed the first separations using GC.[4] Once the validity and utility of the method had been demonstrated, other workers quickly adopted it, and GC became applied more rapidly and broadly to scientific research than any other analytical technique developed before that time. In GC, the sample is vaporized and injected onto chromatographic columns and then separated into many components. The elution is brought about by the flow of an inert gaseous mobile phase. In recent decade the development of gas chromatographic skill and its applications have been significant. Tens of thousands of papers have been published using GC as the analytical technique, and it is estimated that as many as 200,000 gas chromatographs are currently in use worldwide. The global market for GC instruments is estimated to be about $1 billion or over 30,000 instruments annually.[5]

Gas-chromatographic methodology is divided into two classes, depending only on the nature of the stationary phase as the mobile phase is always a gas. These are *gas–solid chromatography* (*GSC*), in which the stationary phase is a solid adsorptive material and solute particles are removed from the mobile phase by electrostatic forces, and *gas–liquid chromatography* (*GLC*), in which the stationary phase is a thin layer of liquid, coated, or bonded on the surface of an inert particle or on the walls of the column itself. In this method solute molecules are retained in the liquid phase based on their partition coefficients between it and the gaseous mobile phase.

Some of the advantages of GC are (a) fast analysis (typically in minutes); (b) efficient, providing high resolution; (c) sensitive, easy detecting ppm and often ppb; (d) non-destructive, making possible on-line coupling, eg, to mass spectrometer; (e) highly accurate quantitative analysis, typical RSDs of 1–5%; (f) requires small samples, typically μL; and (g) reliable and relatively simple and the technique is relatively inexpensive.

The disadvantages of GC include (a) it is limited to volatile samples; (b) not suitable for thermally labile samples; (c) fairly difficult for large, preparative samples; and (d) requires elaborate instrument such as mass spectroscopy, for confirmation of peak identity.[5]

Theory of Gas Chromatography

In the mid-1950s a group of Dutch chemical engineers began a study of the processes that caused band-broadening in chromatography. They derived an expression, commonly called the van Deemter equation, relating the height equivalent to a theoretical plate (HETP) to a number of experimental parameters, including the diameter of stationary phase particles, the diffusion coefficients of the solute in the stationary and mobile phases, and the flow rate of the mobile phase. For descriptive purposes the original, complicated equation frequently is given in the simplified form

$$\text{HETP} = A + B/\mu + C\mu \qquad (12)$$

where μ is the linear velocity, in cm/sec, of the mobile phase, and A, B, and C are coefficients that describe the various diffusion processes occurring in the chromatography that lead to band-broadening.

Coefficient A is called the *eddy diffusion* or *multiple-path coefficient* and is concerned with the different paths traveled by the molecules of a particular solute during their passage through the column. The particles of the stationary phase, whether irregularly or spherically shaped, are packed as tightly as possible, and the solute molecules must pass around them to proceed along the column. Because of the large number of possible paths, some molecules of the same kind will reach the end of the column before others. Faster molecules are found in the leading edge of the peak, and slower ones form the trailing edge. The net effect of this distribution is band-broadening. In a modern chromatographic column, which is packed with small, uniformly sized particles, the value of A is minimal and the contribution of this term to increasing the HETP is negligible. In a capillary GC column, which contains no solid particles, the value of A is zero.

Coefficient B in the van Deemter equation is termed the *coefficient of longitudinal diffusion.* Because the concentration of solute is lower at the edges of the band than in the center, a gradient exists and, during the travel of the band through the column, solute is diffusing continually through the mobile phase away from the center of the band. This phenomenon occurs at both the leading and trailing edges of the peak and contributes further to band-broadening. Because the equation predicts that the contribution to the HETP of this term is inversely proportional to the mobile phase velocity, the effect is more pronounced at low flow rates. Diffusion effects are more severe in GC than in liquid chromatography because diffusion coefficients are several orders of magnitude higher in a gas. The contribution of longitudinal diffusion to band-broadening can be lessened by the proper adjustment of flow rate and by increasing the viscosity of the mobile phase.

Coefficient C, the *coefficient of mass transfer,* is concerned with the transfer of the solute between the two phases. Because the mobile phase is moving rapidly, equilibrium between the two phases may not be attained. Therefore, some solute molecules in the mobile phase are not transferred to the stationary phase quickly enough, and, as a result, are carried ahead of the center of the band. Those in the stationary phase are retained too long and, hence, lag behind. In contrast to longitudinal diffusion, the contribution to the plate height of this term is directly proportional to flow rate; thus, to minimize the overall effect, a compromise in flow rate is necessary. Mass-transfer effects also may be lessened by using a very thin coating of stationary phase so that the area in contact with the mobile phase is maximized while diffusion deep into the stationary phase is reduced.

An efficient GC column will have several thousand theoretical plates, and capillary columns will have in excess of 10,000 theoretical plates. The HETP for a 1 m column with 10.000 theoretical plates would be 100 cm/10,000 plates = 0.01 cm/plate. In an HPLC, efficiency on the order of 400 theoretical plates per centimeter is typically achieved, and columns are 10 to 50 cm in length.

BASIC INSTRUMENTATION

The essential components of a gas chromatograph are the same whether the instrument is an inexpensive student-grade apparatus or a research instrument costing tens of thousands of dollars. The basic components are shown in the block diagram in Figure 33-7.

The *carrier gas,* which serves as the mobile phase, is supplied in steel tanks under high pressure. To reduce the pressure to a level compatible with the requirements of the instrument, a suitable two-stage diaphragm-controlled pressure regulator is fitted to the tank. The carrier gas, now at a pressure of approximately 40 to 80 psi, passes into a flow controller that allows the operator to adjust the flow rate to the desired operating level before the carrier gas moves into the instrument, which is contained within a thermostat controlled chamber capable of achieving temperatures ranging from less than ambient to as high as 400°.
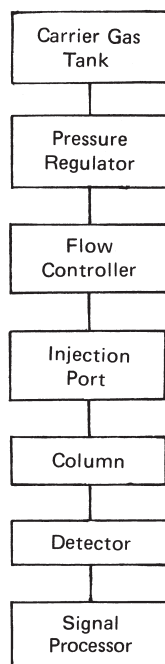
Figure 33-7. Block diagram of a gas chromatograph showing the essential components of the system.

The next component in the line of flow is the sample injection port. This is a small chamber, usually separately heated to a temperature slightly above that of the column, in which the analytical sample is made to vaporize rapidly before entering the column. The sample is introduced into the flowing gas stream through a self-sealing rubber or silicone *septum* using a microliter syringe. Injection of the sample solution may be done either manually or using and automatic injector, which gives more reproducible results. The sample may be injected into the chamber directly on the beginning of the column to minimize diffusion due to turbulence. Samples may be pure liquids, solids dissolved in liquid solvents, or gases. The gaseous mixture next enters the column, which is a tube, usually silica or stainless steel, 1 to 300 m long and with an internal diameter of 0.2 to 4.6 mm. The column may be straight, coiled, or U-shaped.

The interior of the column is filled with either a solid adsorbent material for GSC, or in the case of GLC, a liquid phase coated as a thin layer either directly on the walls or on a packing of small, inert solid particles. Based on their electrostatic attraction for the surface of the solid or their partition coefficients between the two fluids, the solutes are retained temporarily by the stationary phase. As the carrier gas continues to flow, the retained molecules diffuse back into the mobile phase.

At the end of the column, each of the separated solutes exists as a binary mixture with carrier gas and moves into the detector, which also may be heated to a level slightly higher than that of the column to prevent condensation of the solutes. The detector is a device that converts some physical property of the solute, such as thermal conductivity, ionizability, or electron-capturing ability, into an electrical signal that is proportional to the amount of solute in the carrier gas. This is amplified electronically and fed to a suitable signal processor that produces a record of the level of the signal versus time. The output also may be sent simultaneously to a computer for storage and calculation.

CARRIER GAS

The choice of the carrier gas is crucial to the success of the chromatography because it is the mobile phase. In theory any gas may be used, but for practical reasons such as inertness, purity,

availability, and expense, usually helium or nitrogen and occasionally hydrogen, argon, or carbon dioxide are employed. The most prevalent carrier gas in the recent years is helium, and its usage is increased six-fold compared to 1970. One of the most important considerations is the purity of the gas, as a contaminated carrier gas will cause a drifting or elevated baseline or it may deposit its impurities on the column. Trace amounts of water can desorb other column contaminants and produce high detector background. Trace hydrocarbons in the carrier gas can cause a high background with most ionization detectors and thus limit their detectability. Water and trace hydrocarbons can be easily removed by installing a 5Å molecular sieve filter between the gas cylinder and the instrument. Drying tubes are commercially available, or they can be readily made by filling a 6-feet by 0.25″ column with GC grade 5Å molecular sieve. In either case, after two gas cylinders have been used, the sieve should be regenerated by heating to 300° C for 3 hours with a slow flow of dry nitrogen.[5] In addition, it is essential that the carrier gas be inert with respect to the sample components, the column-packing materials, and the components of the instrument. The viscosity also is important because low-viscosity gases such as hydrogen or helium allow higher flow rates to be maintained, whereas a gas with relatively high viscosity, such as nitrogen, may be useful in lessening longitudinal diffusion of the solutes and thereby reducing band-broadening.

STATIONARY PHASE

The interior of a GC column contains either an uncoated solid material for GSC or an inert *solid support* coated with a thin layer of *liquid phase* for GLC. The particles of the packing material are small (80- to 120-mesh) to minimize *void volume* (total volume of interstitial space between particles), while at the same time providing a large surface area for interaction with the solutes. Alternatively, in a column of capillary dimensions (<0.75 mm), the liquid phase may be coated directly on the wall.

In GSC the adsorbents most frequently used are activated charcoal, silica gel, alumina, or glass beads. For the analysis of low-molecular-weight compounds such as water or alcohols, molecular sieves may be used or columns may be chosen from a group of porous polymers made from styrene and divinylbenzene. These are manufactured in such a way that their porosity is controlled carefully, and their separation ability is achieved by a combination of adsorption and size exclusion.

For GLC, the most commonly used solid support material is diatomaceous earth, which is treated with acid and base to remove impurities and then calcined to activate the surface. Nonporous supports such as glass microbeads also have been used. The liquid phase is coated uniformly on the surface of the solid support usually at levels of 1% to 5% by weight. For the separation of compounds that are retained only slightly, amounts as high as 40% have been used. The liquid must be chemically stable, have a low vapor pressure at operating temperatures, and have specific solvent properties toward the compounds to be analyzed.

Today, there are in common use about 15 or 20 highly purified liquid phases that differ from each other in their overall polarity and specific selectivity for particular functional groups on the solute molecules. Most of these are based on silicone polymers with substituents such as phenyl, cyano, or trifluoropropyl introduced to affect polarity and selectivity. Methyl silicone and methyl phenyl silicone are the most commonly used liquid phases, and their use is increased from 28% in 1970 to 74% in 2000. Ethylene glycol polymers also are used frequently, mainly for the separation of polar compounds such as alcohols and amines, but the use of these compounds decreased by almost five times (32% in 1970 to 7% in 2000).[6] In the case of the nonpolar liquid phases, the elution of a mixture of solutes occurs usually in order of increasing molecular weight, as the larger the compound, the more nonpolar it likely is to be and the more strongly it will be retained. As the polarity of the

liquid phase is increased, elution order will be based more on the relative polarities of the solutes, with the most polar substances being retained more strongly.

Several methods have been developed to facilitate the choice of the most-efficient stationary phase for a particular analysis. The Retention Index system of Kovats, a measure of the relative retention of a compound with respect to a series of *n*-alkanes, was formulated to catalog the relative polarities of the liquid phases. If the retention of the compound is determined on a polar and a nonpolar column, the difference in relative retention is a measure of the column polarity.

COLUMN DESIGN

Capillary Columns

Because of the relatively low viscosity of the mobile phase in GC, the contribution of solute diffusion to band-broadening can be substantial. In an effort to reduce the volume within which the solute can diffuse, narrow-diameter columns are used frequently. A small increase in efficiency has been achieved by using tubing with an inner diameter of about 1 mm. However, the greatest increase in efficiency has occurred with the use of *capillary columns*. The column diameter influences (i) broadening of chromatographic zones, (ii) column load quantity per column length unit, and (iii) the retention factor. The preferred column diameter in the capillary GC is 0.25 mm and 0.32 mm. Frequency of use of such columns in published literature is 73%. Researchers are interested in small diameter capillary columns, which are more compact and can be used in express-analysis and for the development of highly efficient columns.[6] The liquid phase is contained within these columns in either of two ways:

In more frequently used *wall-coated open tubular* (*WCOT*) columns the stationary phase is deposited as an extremely thin layer directly on the inner surface of the tube. This may be either as a film or, more frequently, by bonding it chemically to the wall of the capillary column. The latter is advantageous because it prevents *bleeding* or loss of the liquid phase due to its volatility at elevated temperatures.

In *support-coated open tubular* (*SCOT*) columns the inner surface of the tube is coated with a layer of inert support onto which the liquid is coated. Because of the irregularity of the support particles, the surface area of the SCOT column is larger, and therefore a more stationary phase is available to interact with the solutes. However, the mechanics of packing these columns are difficult, and they are not used as often as the WCOT type.

Because of its narrow diameter, the void volume in a capillary column is much lower than in the usual packed column and the ratio of V_S to V_M is high compared to a larger diameter column. In terms of plates per unit length, efficiencies four to five times greater than those of packed columns can be achieved by lessening eddy diffusion and increasing mass transfer (terms *A* and *C* in Equation 12). The absence of particles in the capillary decreases resistance to gas flow and permits the use of columns as long as 300 m. This results in much higher efficiencies, and columns of several hundred thousand theoretical plates are available.

The disadvantage of capillary columns is that they have low capacities due to the small volume of the stationary phase. Therefore, injection volumes must be very small (<0.1 μL) or injections must be made through a splitter that diverts more than 95% of the sample away from the column. The method also is not useful when it is desirable to collect the eluted solutes.

Packed Columns

Large-bore columns with inner diameters of 2.0 to 4.6 mm and packed with inert, solid particles coated with a thin layer of liquid phase also are used. They usually are made of stainless steel or glass and range in length from 1 to 3 m. They are at-tached at either end to the injection port and the detector, using compression fittings to achieve a gas-tight seal. Because the shape of the column has no effect on the chromatographic process, it is designed to conform to the dimensions of the oven. Shorter columns may be straight or U-shaped, but longer ones usually are coiled into a spiral.

For certain compounds, notably steroids, which are highly susceptible to degradation and molecular rearrangement on hot metal surfaces, glass columns are employed widely as their surfaces are relatively inert. The disadvantages of glass are that it is difficult to get a gas-tight seal at the injector and detector connections, and they are brittle enough to break under limited stress.

OPERATING CONDITIONS

Most chromatographic analyses are done in the *isothermal* mode in which the temperature of the instrument is maintained constant throughout the run. However, this method frequently is unsatisfactory for complex mixtures, when both volatile and comparatively nonvolatile solutes are present. If for these mixtures the column is operated at a high temperature, the low-boiling solutes will be eluted rapidly but not resolved while the less-volatile substances may be separated satisfactorily. At a lower operating temperature, all substances may be resolved, but the retention time for the less-volatile compounds will be excessively long and the peaks may be so broad as to be undetectable.

To obviate these problems, the technique of *temperature programming* may be used. In this method the temperature of the column is raised at a preset rate beginning at the time of injection of the sample. The programming rate may be constant during the run or, in more sophisticated instruments, periods of isothermal operation may be interspersed between temperature rises. The result of temperature programming is a chromatogram with evenly spaced peaks having good heights, resulting in an overall savings of time. The initial temperature should be chosen to minimize the retention time for the least-retained solute; the final temperature must be sufficient to elute the least-volatile compound in a reasonable time without exceeding the operating limits of the liquid phase.

DETECTORS

Detectors identify solutes as they exit the chromatographic column. As solutes are eluted from the GC column, they interact with the detector. The GC detector converts this interaction into an electrical signal that is sent to the data system. The magnitude of the signal is plotted versus time, and a chromatogram is generated. Gas chromatography detectors use one of several technology types to identify solutes as they exit the column.

Some of the commonly used methods include flame ionization, thermal conductivity, electron capture, nitrogen-phosphorous, flame photometric, and photo ionization type detectors as summarized in Table 33-1. These are grouped into two general classifications, *mass flow rate* detectors, which are sensitive to the rate of flow of the solute through the detector, and *concentration-sensitive* detectors, which respond to the concentration of solute in the mobile phase in the detector.[6,7]

The *Thermal Conductivity Detector* (*TCD*), also called the *Hot Wire Detector* (*HWD*) or *katharometer*, is a concentration-sensitive detector and is used widely because it is a universal detector in that it responds to all solutes. In this device, a coil of fine wire, usually made of a tungsten–rhenium alloy, resides in a small chamber into which the column effluent flows. In practice most TCDs consist of a matched pair of wires, one of which is placed in the gas stream before it enters the column, while the other is at the end of the column. An electrical potential is placed across the wire filaments and they heat up due to their resistance.

Table 33-1. Summary of Commonly Used GC Detectors

NAME OF DETECTOR	TYPE	CARRIER GASES	SELECTIVITY	MINIMUM DETECTIBLE QUANTITY	OTHER CHARACTERISTICS
Flame ionization (FID)	Mass flow	Nitrogen or Helium	Organic compounds only	10^{-11}g ($\sim$50 ppb)	Excellent linearity and stability, Temperature limit $-400°C$
Thermal conductivity (TCD)	Concentration	Helium	Universal (all compounds)	10^{-9}g ($\sim$10 ppm)	Excellent linearity and good stability, Temperature limit $-400°C$
Electron capture (ECD)	Concentration	Make-up	Halides, nitrates, nitriles, peroxides, anhydrides, organometallics	10^{-9} to 10^{-12}g	Good linearity and fair stability
Nitrogen-phosphorus	Mass flow	Hydrogen and air	Nitrogen, phosphorus	10^{-12}g for P and 10^{-11}g for N_2	Very sensitive and highly specific
Flame photometric (FPD)	Mass flow	Hydrogen and air possibly oxygen	Sulphur, phosphorus, tin, boron, arsenic, germanium, selenium, chromium	10^{-7}g	Quenching or re-absorption of the light emitted by the selected species
Photo-ionization (PID)	Concentration	Make-up	Aliphatics, aromatics, ketones, esters, aldehydes, amines, heterocyclics, organ osulphurs, some organometallics	20^{-9}g	Good linearity and stability

Data from McNair HM, Miller JM. *Basic Gas Chromatography.* New York: John Wiley, 1997.

The resistance of the filament is a function of its temperature. When only carrier gas is flowing through the chamber, the filaments maintain a steady temperature, which is determined by the thermal conductivity of the gas; however, when a binary mixture of solute and carrier gas emerges from the column, the mixture has a different thermal conductivity and heat is conducted from the sample filament at a greater or lesser rate. This changes the resistance of the wire, and the change in resistance or current is a measure of the concentration of the solute in the detector.

The thermal conductivities of hydrogen and helium are as much as 10 times higher than those of most organic compounds, so even a small amount of solute will cause a large change in the output of the detector. Nitrogen, however, has conductivity close to that of most organic compounds, so sensitivity is lower with this carrier gas and, in fact, it is possible to obtain negative peaks. TCDs are simple, inexpensive, and nondestructive to the sample. They are relatively insensitive, however, compared to other detectors, and are generally not useful for analyses requiring detection of low levels of solutes such as drugs in biological fluids.

The *Flame Ionization Detector* (*FID*) is a mass flow detector and is the most frequently used detector in GC because it is highly sensitive, able to detect microgram quantities of solutes, and an almost universal detector. It responds well to most organic compounds but is insensitive to water and most inorganic substances. In this device, hydrogen and air or oxygen are introduced into the column effluent stream. The mixture is ignited and, as a result of the energy of the flame, electrons are stripped from the solutes and ions are formed. These charged particles migrate to a pair of oppositely charged collector electrodes in the chamber and cause a small electrical current to flow. The current, which is amplified to produce a useful signal, is proportional to the rate of flow of solute through the detector. In addition to their exceptional sensitivity, flame-ionization detectors are useful because of their large *linear dynamic range*. They will respond in a linear manner to amounts of solute that differ in concentration by several orders of magnitude.

Other highly sensitive detectors are available but their response is limited to compounds containing certain specific functional groups, such as nitrogen, phosphorus, or the halogens. This property of selectivity can be very useful, but even if the chromatographic separation is not optimal, an interfering solute will not be detected unless it contains the functional group for which the detector is specific. Thus, they impart a second level of selectivity to the procedure.

The *Electron Capture Detector* (*ECD*) is one of the most sensitive of detectors, being able to respond to nanogram, or even picogram, quantities of materials having functional groups that possess high electron affinity, such as the halogens or nitro groups. In this device, a radioactive source, usually ^{63}Ni, emits beta particles that interact with the carrier-gas molecules to form positive ions and electrons. These, in turn, migrate to oppositely charged electrodes in the detector chamber to produce a *standing current*. When a detectable solute elutes from the column, it is able to "capture" some portion of the electrons, thereby lowering the standing current. This decrease in the current is detected electronically and is proportional to the amount of solute.

The ECD is extremely sensitive to low levels of halogenated compounds such as pesticides. However, its linear response range is narrow, and it is very susceptible to permanent saturation if it is exposed to too high a concentration of a halogenated compound.

The *Thermionic Specific Detector* (*TSD*), also called the *Nitrogen Phosphorus Detector* (*NPD*), is a modified form of the FID that shows increased response to compounds containing nitrogen and phosphorus. It consists of a standard FID with an electrically heated bead of a solid alkali metal compound, such as rubidium silicate, suspended in the area above the flame assembly. In the presence of excess air, a plasma is formed in the area of the bead. This produces large numbers of ions from nitrogen- and phosphorus-containing compounds that are then detected at the collector electrodes, as in the FID. The mechanism of action is not understood fully, but its sensitivity for nitrogen- and phosphorus-containing compounds is 10^3 to 10^4 times greater than for other organic compounds. It, too, has important applications in pesticide-residue analysis.

Another modification of the FID, which has increased selectivity for sulfur and phosphorus containing compounds, is the *Flame Photometric Detector* (*FPD*). The eluted compounds first are burned in the usual FID flame, from which the products of the pyrolysis then pass to another flame where sulfur and phosphorus atoms are excited to a higher energy state and subsequently detected by emission spectroscopy. The sensitivity of this device to sulfur and phosphorus is about 10^5 times greater than that for carbon compounds.

Probably the most sensitive and useful analyses can be made by combining a gas chromatograph with a mass spectrometer. Using a suitable separator to remove the carrier gas allows direct introduction of the solute into the ionization chamber of the *Mass Spectral Detector* (*MSD*) after it exits the

column. This technique has the advantage of extremely high sensitivity (10^{-12} to 10^{-15} g) so that usually only one injection of the unknown is required. The GC-MS technique also is useful for quantitative analysis by using selective-ion monitoring.

There are a number of other detectors that are available for use in GC but are used less frequently because they do not offer significant advantages over the devices currently employed. These include detectors whose operating principles are based on coulometry, conductivity, or photoionization.

Some novel detector designs were reported during the past years including a chlorine-selective pulsed discharge emission detector (CI-PDED). This detector is based on reaction of krypton with chlorine and a unique detector design. A krypton ion produced in the krypton-doped helium pulsed discharge reacts with chlorinated compounds within the pulsed discharge to produce an excited species of KrCl* which emits at 221-222 nm[7].

SPECIAL TECHNIQUES

It is not unusual in the practice of GC to encounter samples that cannot be analyzed satisfactorily no matter what combination of mobile and stationary phases are used. For example, petroleum fractions contain tars and other high-boiling hydrocarbons that chromatograph with difficulty, if at all. In addition, many drugs that contain carboxylic acid or primary amine functional groups are volatile enough to chromatograph but will give badly tailed peaks due to nonideal interactions of the functional groups with the stationary phase. However, for research and quality-control purposes, the pharmaceutical and chemical industries require that these substances be analyzed; to overcome the problem posed by these compounds, special techniques such as *pyrolysis* and *derivatization* have been developed.

Pyrolysis GC is used frequently for the analysis of very high-molecular-weight compounds such as crude oil fractions, rubber vial closures, and packaging materials. In this technique, the high-molecular-weight substances are decomposed to lighter and more volatile compounds by controlled heating in a furnace, which may be external to the gas chromatograph or an integral part of the instrument. The resulting lighter compounds then will be chromatographed as usual, frequently using capillary columns.

Because the nature of the decomposition products rarely is known with any certainty, the chromatogram that is produced represents a "fingerprint" of the original sample. If the time and temperature of pyrolysis are controlled carefully, the method is reproducible and valuable for checking raw materials from different suppliers to determine if the source or the chemical composition has changed over a period of time.

A number of compounds, such as steroids, do not chromatograph well because they are not sufficiently volatile or they decompose at the higher temperatures needed for successful GC. Others, such as fatty acids, yield poorly shaped peaks. It frequently is possible to obviate these problems and obtain good chromatograms by forming derivatives of these substances. Many of the procedures used to produce derivatives in these cases are the same as those used in qualitative organic analysis, such as acylation of alcohols, formation of oximes and hydrazones of carbonyls, or esterification of fatty acids. However, for GC, a different class of derivatizing reagents called *silylating agents* has been used most often. These agents are intended to react with compounds containing labile protons such as alcohols, amines, carboxylic acids, or thiols to produce the corresponding ethers, silyl amines, esters, or thioethers. Because of the reduced polarity, the derivatives have greater volatility and stability than the original compounds. A number of derivatizing agents have been used, the most common of which are *N*-trimethylsilylimidazole (*TSIM*), *N,O*-bistrimethylsilyltrifluoroacetamide (*BSTFA*), and *N,O*-bistrimethylsilylacetamide (*BSA*). The by-products of the reactions are very volatile and elute very rapidly so they do not interfere with the chromatography.

Other than being useful as derivatizing agents, silylating compounds also are used to deactivate solid supports and the surfaces of glass columns. In these applications, they react readily with the silanol groups on the silica surface, thereby blocking polar sites that would interfere with the separation process.

QUALITATIVE ANALYSIS

Although GC is used primarily as a quantitative technique, it also is valuable in the qualitative analysis of unknown substances. This may be accomplished in either of two ways: by comparing the retention parameters of the unknown with known compounds, or by trapping the effluents as they leave the column and subjecting them to classic chemical or spectrometric identification procedures.

In terms of parameters derived from the chromatogram, the retention or relative retention times or their corresponding volumes are useful indicators of identity when compared with the same parameters for a known compound. The related variables, the capacity factor, k', and the separation factor, α, also may be used. If the unknown compound is suspected of being a member of a homologous series and sufficient known members of the series are available, plots of $\log t_R$ versus carbon number or $\log t_R$ on a polar column versus $\log t_R$ on a nonpolar column will give a straight line for homologs.

It also is possible to identify solutes after GC by collecting individual fractions as they elute from the column. This can be done manually or automatically using a fraction collector that is activated by the signal from the detector. In this manner the entire procedure can be automated and carried out unattended over an extended period of time to ensure that adequate amounts are collected for subsequent analysis.

QUANTITATIVE ANALYSIS

In addition to providing rapid and efficient separations of complex mixtures and qualitative information about the eluted substances, GC also can furnish the analyst with accurate and precise quantitative data.

The parameter that is proportional to the concentration of a compound in the GC effluent is the area under the elution peak, which is the integral of the elution curve from the point where it leaves the baseline to the point where it returns. Using computerized techniques, this integral can be determined exactly; however, a number of manual integration methods may be employed. These are based on the assumption that the shape of the peak is Gaussian; although they do not yield the true peak area, the results obtained are proportional to it and may be used with equal confidence. Some of the methods used for integration are

Triangulation—Tangents to the inflection points of the peak are drawn from the baseline to the point where they meet above the peak. The third side of the triangle is drawn along the baseline, and the area is determined by multiplying the base width by one-half the height. The resulting value is equal to 96% of the actual area of the Gaussian peak.

Height Times Width at Half-Height—The baseline portions of the chromatogram before and after the peak are joined with a straight line. The height is measured from this base and multiplied by the width at one-half height. The result is equal to 84% of the true area.

Height—If the conditions of temperature and flow rate are controlled rigorously, the peak height produced by a given quantity of solute will be constant from run to run. It therefore may be used directly as an estimate of the area.

Computer Integration—The computer, or integrator, converts the analog voltage produced by the detector into a digitized quantity and computes the results. This is by far the most precise method, as well as being applicable to peaks whose shape is not ideally Gaussian.

Once the relative areas of the peaks in the chromatogram have been determined, these data, which are proportional to the concentration of each of the species, must be used to determine exact concentrations. This is accomplished in one of three ways:

Area Normalization—The assumption is made that each substance in the injected mixture produces a separate peak in the chromatogram. The weight of the material in any peak is found by

determining the ratio of its peak area to the sum of the total areas of all the peaks and multiplying this by the total weight of solute in the amount injected. This method is used infrequently because the initial assumption usually is not valid. It is, however, the standard default method on computerized integrators and is useful as a check on that reproducibility of repeated injections of the same sample.

External Standardization—A pure reference standard material corresponding to the substance to be determined is dissolved in a solvent at a known concentration. Exactly measured quantities of this solution (1, 2, 3, 4, and 5 μL) are injected successively. The areas of each of the peaks produced are plotted versus the mass of solute injected and a calibration curve is produced. Next, the unknown solution is injected, the area of the peak determined, and the concentration found by interpolation. This method is very accurate, but it is necessary that the analyst be skilled in the use of microliter syringes or that an automatic injector be used because the volumes injected must be known exactly. It also is assumed that instrument parameters remain constant for the period during which the samples are introduced, an assumption that is not always valid.

Internal Standardization—To obviate the difficulty of introducing precisely measured quantities into the GC, the internal standard method may be used. This procedure requires two standards: the *analytical standard*—a pure sample of the compound to be analyzed, and the other—an *internal standard*. This is normally a substance that elutes at a position near the substance being analyzed and is well resolved ($R_S > 1.25$), but that cannot be converted to the analyte under the conditions of the analysis. A series of solutions is prepared containing varying amounts of the analytical standard and constant amounts of the internal standard. These are chromatographed and a calibration curve is determined by plotting the ratio of the areas of the two peaks versus the ratios of their concentrations. The unknown then is dissolved in a suitable solvent, the same amount of internal standard is added, and the mixture is chromatographed. The ratio of the areas is calculated and, by interpolation on the calibration curve, the amount of the unknown determined. This method is the most frequently used technique for quantitative analysis by GC because it is not necessary to know the exact amount of solution injected. Usual practice is to prepare a solution of the internal standard in the solvent employed to dissolve reference standards or samples, thus ensuring unvarying ratios of standard or sample peak areas to that of the internal standard.

LIQUID CHROMATOGRAPHY

In 1906 Michael Tswett published a comprehensive paper on liquid chromatography in which he clearly explained the nature of the process and his appreciation of its potential. The method was not adopted widely until many years later. In 1941 Martin and Synge,[3] who had been unsuccessful in using countercurrent extraction for the separation of amino acids in wool samples, developed a liquid-chromatographic process in which they used a packed column containing water-saturated silica gel and a mobile phase of butanol-chloroform. They perfected the experimental techniques and explained the theoretical aspects of the procedure so thoroughly that they were awarded the Nobel Prize for this work in 1952.

Since that time, liquid chromatography has become one of the most versatile techniques available to the analyst because of its simplicity and capacity for high-resolution separations. Separations may be developed based on such diverse characteristics as the polarity of the solutes, their ionic nature, their molecular weight, their partitioning ability, or their ability to form affinity complexes.

The term liquid chromatography is used today to refer to those methods in which the separation takes place within a packed column. The packing material is the stationary phase and may be a solid with adsorptive or exclusion capabilities or an inert support coated with a liquid phase. A liquid mobile phase is used as the eluant. Although thin-layer chromatography and paper chromatography use a liquid mobile phase and a solid stationary phase, they differ in that the separations take place on a planar surface rather than in a column.

Liquid chromatography can be performed using either of two methods:

1. The classic procedure developed by Tswett, called *open-column chromatography*, in which the mobile phase is allowed to flow through the packed column under the influence of gravity or, at most, low pressure (eg, 50–100 psi).

2. The procedure in which the mobile phase is forced through the packed column under high pressure. The latter method is called *high-performance liquid chromatography* (*HPLC*) because of the extremely high efficiencies (as many as 50,000 plates/m) attainable, or *high-pressure liquid chromatography* because of the high pressures (1000–3000 psi) required. In HPLC, particle diameter is typically 10 μm or less and, as a result, columns are packed more tightly and develop high back pressures that necessitate pumping the mobile phase through the column.

Whether HPLC or open-column methods are used, the mode of separation depends primarily on the nature of the stationary phase. Five modes are available: adsorption, partition, ion-exchange, size-exclusion, or affinity. Each of these will be discussed in detail.

ADSORPTION CHROMATOGRAPHY

In the adsorption mode of liquid chromatography, as in gas–solid chromatography, solutes are retained as a result of the ability of the stationary phase to bond them temporarily to its active surface. The forces involved usually are relatively weak and effective only over short distances. These include van der Waals and London forces, dipole and induced-dipole interactions with polar groups on the active surface, charge-transfer forces, and hydrogen-bonding.

With this type of binding, termed *physical adsorption,* the energy required to break the bonds is small, and the mobile phase, through its ability to dissolve and displace the solutes effectively, can counteract these attractive forces. Therefore, an efficient chromatographic process can occur based on the competition between dissolution of the solute in the mobile phase and binding at the surface of the stationary phase. However, when stronger chemical bonds form between solutes and the adsorbent, as in the process of *chemisorption,* the mobile phase is not able to provide sufficient energy to desorb the solutes. In this case, equilibrium between the two phases is not reached, and solutes are adsorbed irreversibly or give unsatisfactory, tailed elution peaks.

Some of the most common of the large variety of substances that have been used as adsorbents are shown in Table 33-2. In addition to adsorptivity, surface area, particle size, and surface activity are of primary importance in determining the utility of a potential adsorbent. A large surface area is necessary to provide effective contact between the two phases and ensure frequent exchange of the solute. Although areas of 5 to 200 m²/g are quoted by suppliers of adsorbents, these values may be lower than the actual effective surface area, because the methods of measurement used do not account accurately for the porous nature of the adsorbent particles or the true shape of the solute molecules. Particle size is important, not only as an indicator of surface area, but also because it determines the resistance of the packed column to solvent flow. Although very small particles may provide a large area for solute interaction, they may pack so tightly that a reasonable flow rate cannot be achieved without using high-pressure techniques with a

Table 33-2. Adsorbents Used in Column Chromatography

Sucrose	(weakest)
Starch	
Inulin	
Talc	
Calcium carbonate	
Calcium phosphate	
Magnesia	
Silica gel	
Magnesium silicate	
Alumina	
Charcoal	(strongest)

Table 33-3. Characteristics of Solvents Used in Chromatography

SOLVENT	ELUOTROPIC VALUE, E^0	DIELECTRIC CONSTANT	SOLUBILITY PARAMETER
Heptane	0.00	1.92	7.4
Hexane	0.01	1.88	7.3
Isoctane	0.01	1.94	7.0
Cyclohexane	0.04	2.02	8.2
Carbon tetrachloride	0.18	2.24	8.6
Toluene	0.29	2.38	8.9
Benzene	0.32	2.27	9.2
Ethyl ether	0.38	4.33	7.4
Chloroform	0.40	4.81	9.1
Methylene chloride	0.42	8.93	9.6
Tetrahydrofuran	0.45	7.58	9.1
Acetone	0.56	20.7	9.4
Dioxane	0.56	2.25	9.8
Ethyl acetate	0.58	6.02	8.6
Acetonitrile	0.65	37.50	11.8
Pyridine	0.71	12.30	10.4
l-Propanol	0.82	20.33	10.2
Ethanol	0.88	24.30	11.2
Methanol	0.95	32.70	12.9
Acetic acid	large	6.15	12.4
Water	large	78.54	21.0

consequent loss in sample capacity. Particles in the range of 75 to 150 μm in diameter are a useful compromise for open-column chromatography, providing a large surface area with good permeability. Surface activity refers to the energy of the active site of the adsorbent, and may vary depending on the nature of the substance and on the amount of water adsorbed. To provide a reproducible surface it is common practice to activate an adsorbent by heating it to expel most of the water and then to deactivate it to a desired level by exposure to a climate of known humidity, returning a known quantity of moisture to the adsorbent.

Among the commonly used adsorbents, silica gel and alumina have surfaces rich in hydroxyl groups and oxygen atoms, thus they interact strongly with polar solutes. Charcoal, activated at 1000° to make it nonreactive to polar compounds, has a very porous surface that slows down the adsorption–desorption process and makes it more prone to chemisorption. Separations on charcoal are based mainly on molecular weight, with larger compounds being retained more strongly. Magnesium silicate has an acid surface characteristic of the insoluble silicates and is similar to alumina in adsorptive properties.

Table 33-3 lists a number of the solvents most commonly used in liquid–solid chromatography cataloged in a standard order according to their relative energy of adsorption per unit surface area on alumina. A listing such as this is called an *eluotropic series,* and, although the relative energies of adsorption differ slightly on other surfaces, the choice of solvents usually is made according to this series. An exception to this is made in the case of charcoal; because of its tendency to adsorb nonpolar substances, the order of solvent strength is reversed.

The solvent used for a particular separation must be chosen with regard to the properties of the solutes as well as the stationary phase. For example, if a group of very polar compounds is to be separated using silica gel, the solvent must be polar enough to overcome the strong attraction between the solutes and the surface or very large retention times will result. If a mixture of less-polar solutes is to be analyzed, a weaker solvent must be used to permit a longer residence time on the column and more equilibrations between the phases. For a more detailed discussion of the methods used to correlate solute structure with retention time using adsorption chromatography the text by Snyder (see *Bibliography*) is recommended.

PARTITION CHROMATOGRAPHY

In this mode of liquid chromatography, mixtures of solutes are separated according to the relative tendencies of their components to partition between a mobile phase and a stationary phase consisting of a layer of liquid coated or bonded onto the surface of a solid support. The liquid is present as an extremely thin layer so that equilibration between the phases may be attained rapidly by minimizing the diffusion of the solutes into the stationary phase. The surface of the solid support frequently is treated (eg, by silylation) to eliminate adsorptive effects.

Although it has been used for many successful analyses, liquid–liquid partition chromatography was, until relatively recently, an inconvenient method to use experimentally. The liquid phase had to be coated onto the solid support by evaporation of a solution or by injecting it onto the column with the mobile phase flowing. In either case, it was difficult to obtain stationary phases that were stable and reproducible. In addition, the choice of mobile phase necessarily was restricted to those in which the liquid coating had limited solubility. For example, if a polyethylene glycol was to be used to provide a very polar stationary phase, a mobile phase of hexane or some other hydrocarbon of very low polarity had to be used. Even so, the liquid stationary phase would be stripped slowly but continually from the column, thereby changing the characteristics of the separation. To prevent stripping, either the mobile phase was saturated with the liquid phase material, or a precolumn containing a high concentration of the liquid phase coated on a solid support was inserted into the system before the analytical column. In either case, the nature of the mobile phase was changed and partition coefficients became less favorable.

Recently, the problems presented by unstable stationary phases have been solved by the development of *bonded-phase chromatography,* in which the liquid phase permanently is bonded chemically to the surface of the solid support. Silica gel, with its high surface population of hydroxyl groups, provides an excellent medium onto which various substances can be bonded using appropriately substituted silylating agents. For example, octadecyldimethylsilyl chloride reacts with silica gel to form a stable, nonpolar stationary phase called ODS (octadecylsilyl). Because of steric effects, not all of the hydroxyl groups of the silica gel are derivatized by the ODS reagent, so the remainder then are reacted with trimethylsilyl chloride in a process called *capping* (or *end-capping*) to reduce adsorption effects. Bonded phases are advantageous in that they can be made reproducibly from batch to batch and the surface does not change during the chromatographic process. They have the disadvantages of being expensive and effective only over the pH range within which the backbone of silica gel is stable, usually pH 2 to 7. Compared to the inconveniences of the former method, however, these disadvantages are not very restrictive, and the development of newer bonded phases using a polymeric support that is stable over a pH range of 1 to 13 promises to alleviate this problem.

Partition chromatography may be conducted in either of two ways: *normal* or *reversed phase.* In the normal phase mode, the stationary phase is a polar substance, such as polyethylene glycol or the untreated silica surface itself, and the mobile phase is nonpolar (eg, hexane). Under these circumstances polar compounds are retarded preferentially and nonpolar substances elute more quickly. In reversed-phase chromatography the stationary phase is nonpolar (eg, ODS) and the mobile phase is polar, usually a mixture of water, methanol, and/or acetonitrile. Nonpolar compounds are retained more strongly by this system, while polar solutes elute first. Reversed-phase separations are the most frequently used methods in HPLC.

Because of the efficiency and availability of reversed-phase materials, especially the ODS or C-18 type, attempts have been made to use them to separate mixtures of ionic compounds such as amino acids. Normally these compounds would not be retained in a reversed-phase packing, because they are too polar to partition appreciably onto the nonpolar stationary phase. Several techniques, all of which involve altering the mobile phase, have been developed to permit successful chromatogra-

phy of ionic compounds using these stationary phases. These methods are called *ion-suppression chromatography, ion- pairing chromatography,* and *"soap" chromatography*.

Ion suppression is used for substances such as weak acids (pK_a >2) and weak bases (pK_a <8), which only are ionized partially at the neutral pH values characteristic of the usual mobile phases. For example, a carboxylic acid with pK_a = 5 will, at pH = 7, be present in both ionized and un-ionized forms with the anionic carboxylate predominating by a ratio of 100 to 1. To enhance the retention of the substance in a reversed-phase system, the pH of the mobile phase can be adjusted to a value low enough to suppress the ionization of the acid, for example, pH < 3. This causes the free acid to predominate and, as it is much less polar than the anion, it will be able to partition into the stationary phase.

For stronger acids or bases that remain ionized throughout the pH range (2–7) where silica is stable, ion-pairing chromatography is the technique of choice. In this method a reagent that dissociates to give ions opposite in charge to those of the solutes is added to the mobile phase. Although the mechanism of action has not been explained fully yet, the added ions may interact with the charged solutes in two ways.

First, they may combine directly with the charged solutes to form ion pairs that are nonpolar and will partition more readily into the stationary phase.

Alternatively, the nonpolar end of the ion-pairing reagent may itself partition into the stationary phase, leaving its polar end extending from the surface into the mobile phase, where it acts as an ion exchanger.

In either case, the retention of ionic solutes in the reversed phase materials is increased. Examples of ion-pairing reagents are heptanesulfonic acid, used for cationic species such as protonated amines, and tetra-*n*-butylammonium hydroxide, which pairs with anionic substances.

The third method, *soap chromatography,* is actually a form of ion-pairing in which the added reagent is a detergent or soap. Examples are sodium lauryl sulfate for cations and cetyltrimethylammonium chloride for anions. Soap chromatography is useful especially for the separation of proteins, because the soap not only neutralizes the charge on the molecule but also affects the conformation of the protein to allow it to interact more favorably with the stationary phase. The practical aspects of ion-pair chromatography are discussed in more detail by Gloor and Johnson.[8]

Another special technique that can be used with partition chromatography is *metal-ion complexation.* In this process, a small quantity of a metal ion, such as Ag^+, is added to the mobile phase in the chromatography of olefinic compounds. The ionic silver interacts with the double bonds, forming charge-transfer complexes and altering the partitioning behavior of the olefinic solute. This technique is useful for separating mixtures of compounds that differ in the extent and placement of the unsaturation.

Centrifugal partition chromatography (CPC) is a combination of countercurrent chromatography and partition chromatography where an automated liquid-liquid extraction process permitting hundreds of successive extractions taken place. The solute equilibrates between the stationary and mobile liquids by the phenomenon of partition. CPC is unique because no solid support is used for the stationary phase. Instead, the liquid stationary phase is retained in the column by a combination of centrifugal force, the special column geometry and the density difference between the two liquid phases.[9]

When a mixture of components is introduced into the mobile phase of the CPC column, it distributes according to the individual components' distribution coefficients while passing through the column. The centrifugal force field applied to the coiled columns promotes the retention of the stationary phase against a continuous flow of mobile phase. The mobile phase flow enables the two phases interact sufficiently for partition to occur.

CPC is used for many types of separations; racemic mixtures, natural products and amino acids are commonly separated on centrifugal partition columns. CPC is also utilized for enzymatic reaction. Some reviews have been published describing the basic principles of CPC and its applications.[10–12]

ION-EXCHANGE CHROMATOGRAPHY

Although ion-pairing techniques have proved useful in many cases for the separation of mixtures of ionic substances, the usual method for the analysis of these compounds is *ion-exchange chromatography*. This method provides a greater degree of selectivity due to the larger number of combinations of mobile and stationary phases that can be employed. It is especially useful for inorganic cations, amino acids, or similar groups of closely related compounds.

The stationary phase materials used to effect these separations are called *ion exchangers,* and they comprise a group of natural or synthetic organic or inorganic polymers that are capable of reversibly removing ions from a solution, while at the same time replacing them with ions of equivalent charge. At all times during this exchange process the principle of electro -neutrality must be obeyed both in the ion exchanger and the solution. An ion exchanger contains *fixed ions,* which are incorporated permanently into its insoluble skeleton, and loosely bound *counterions*, which are opposite in charge to the fixed ions and capable of being exchanged when charged species are adsorbed from solution. If the counterions are charged positively, the material is called a *cation exchanger;* if negative, it is an *anion exchanger.*

The inorganic polymers used in this type of chromatography are aluminosilicates, which have lattice, or cage-like, structures. Because of the preponderance of oxygen atoms in the polymer, it is charged negatively and the counterions, usually calcium or sodium, are positive. Therefore, they are cation exchangers. The naturally occurring members of this group are called *zeolites,* while the synthetic ones, which were developed to provide standardized structures with constant pore sizes, are called *molecular sieves.* Because of their low capacities for ion exchange, these inorganic substances are used primarily for the size separation of small molecules.

The most frequently used ion-exchange materials are organic copolymers made from styrene (vinylbenzene) and divinylbenzene (DVB). The styrene polymerizes to give long, twisted chains of carbon atoms, with a benzene ring at every other carbon. Divinylbenzene is added to cross-link these chains and give a three-dimensional bead-like structure. Commercially available ion-exchange resins are identified according to their percent cross-linking, as ×2, ×4, ×6, and so on, corresponding to the initial percentage of DVB in the reaction mixture.

Because the styrene–DVB copolymers have no intrinsic ion-exchanging properties of their own and act only as a skeleton, charged functional groups must be added. Reaction with chlorosulfonic acid places a sulfonic acid group on each of the non-linked benzene rings, yielding a *strong cation exchanger,* that is, one in which the counterions can be removed easily from the fixed ions. If methacrylic acid is used in the polymerization in place of styrene, the resulting copolymer has carboxylic acid groups attached to the skeleton and functions as a *weak cation exchanger*, that is, one in which the counterions do not dissociate at low pH. *Strong anion-exchange* resins can be made from the same skeleton by introducing quaternary amine functional groups, while *weak anion exchangers* use polyamines as the ionizable groups.

Many other substances are used both as the skeletal components and the functional groups for ion exchange. Carbohydrate polymers, such as dextran and cellulose, when used as the insoluble matrix change the selectivity of a resin. For example, solute ions with attached polyaromatic groups, such as the anthraquinonesulfonic acids, do not chromatograph well on polystyrene-based ion exchangers, because they associate too strongly with the benzene rings of the resin. On a cellulose-based exchanger, however, separation is possible because the mechanism is limited entirely to the ion-exchange process. Silica gel also is used as a support matrix for preparing ion exchangers, especially in HPLC, where strength of the particle is important, as it must not be crushed by the high operating pressure of the system.

Other functional groups used frequently are diethylaminoethyl (DEAE) and triethylaminoethyl (TEAE), both of which are anion exchangers, and carboxymethyl (CM), which is used in cation-exchange resins. Attached to matrices of cellulose or dextran, these substances have been employed widely for the separation of proteins and peptides.

The mechanism of action in this mode of chromatography depends on the replacement of the counterions of the resin by the ionic species being separated. This can be illustrated by the procedure used for purifying water by passing it through a mixed-bed resin. Using sodium chloride as a typical contaminant, the mechanism is

(1) $RESIN—SO_3H + Na^+ = RESIN—SO_3Na + H^+$

(2) $RESIN—N(CH_3)_3OH + Cl^-$ (13)

$$= RESIN—N(CH_3)_3Cl + OH^-$$
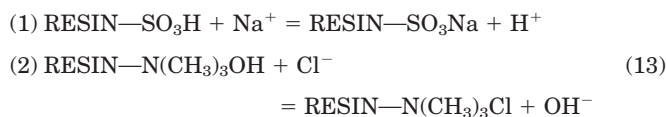
Water of the exceptionally high purity needed for making mobile phases for HPLC is prepared by an ion-exchange column, followed by passage through a charcoal adsorption column to remove nonionizable organic compounds and then microfiltrated to exclude particulate matter and bacteria.

For a mixture of solute ions that differ in charge, the more highly charged species are retained preferentially. Thus, on a sulfonic acid resin, aluminum is bound more strongly than calcium, and calcium more strongly than sodium. The binding of negatively charged species to strong anion exchange resins follows the same trend. Among substances of the same charge, retention is related to the size of the hydrated ions, with smaller ions being held more tightly. Because the smaller elements in the periodic table bind more molecules of water, their hydrated ions are larger; therefore, for the alkali metals the order of retention is $Cs^+ > Rb^+ > K^+ > Na^+ > Li^+$.

Another parameter that affects retention is the nature of the substituents attached to the charged portion of the solute species. Polystyrene resins exhibit preference for ions containing aromatic groups over aliphatic groups, because in addition to binding due to the electrostatic forces of ion exchange, the aromatic groups of the solute interact directly with the skeleton of the resin.

Ion-exchange chromatograms may be developed either by displacement or by elution methods. In the former case, an ion that is retained more strongly than any of the solute ions displaces them from the resin, and a continuous series of bands results (see Fig 33-3). In elution development, the eluting agent is an ion for which the resin has less selectivity than it has for the solute ions. Transfer of the solute ions to and from the resin depends on their exchange equilibria with the eluting ion. The resulting chromatogram consists of a series of separate Gaussian peaks, as in Figure 33-4.

Mobile phases used in ion-exchange chromatography are usually aqueous salt solutions that may be buffered to a desired pH or adjusted to a constant ionic strength. The choice of the mobile phase depends on knowledge of the selectivity of the resin for the solute ions and the influence of solution equilibria due to pH or complexation. Mixed aqueous-organic or organic solvents may be used if the stationary phase is not altered. Gradient elution is used for difficult separations.

Chromatofocusing, first described in 1978 by Sluyterman and Elgersma,[13] is a special method of ion-exchange chromatography that is of great utility in the separation of mixtures of proteins. In this case a buffer, adjusted to a specific pH, is added to an anion-exchange column previously adjusted to a different pH. As the buffers mix, a pH gradient is formed along the length of the column, ranging from the initial pH at the far end to that of the added buffer at the beginning. If the pH at the start of the column is lower than the isoelectric point of the protein to be analyzed, it will carry a positive charge and will not interact with the anion exchanger. Instead, it will migrate along the column to a point where the pH is just greater than the isoelectric point, at which time it will acquire a negative charge and bind to the resin. Thus, a group of proteins will arrange themselves on the column in order of their isoelectric points. As the pH gradient moves down the column, the proteins will migrate downward, so as to remain negatively charged, until each elutes from the column at its isoelectric point. The fractionation of complex mixtures of proteins is therefore possible using this method of separation.

SIZE-EXCLUSION CHROMATOGRAPHY

Size-exclusion chromatography (SEC), also called *gel filtration or molecular-sieve chromatography,* is an efficient technique used to separate groups of solutes based on their effective size in solution. The stationary phases used to attain these separations are polymers that have been cross-linked to yield an open network with numerous pores of consistent size. The degree of cross-linking is controlled carefully to yield a series of gels having different pore sizes and fractionation ranges. When a mobile phase containing a mixture of solutes of various sizes is passed through a column of these materials, molecules that are too large to fit within the pores are "excluded" and remain completely in the mobile phase. They are, therefore, eluted rapidly near the void volume. Molecules of smaller size are free to diffuse in and out of the pores so that, in effect, their path through the column is longer and they will elute later, as is depicted in Figure 33-8. The extent of retention depends on the size of the included molecules relative to the size of the pores. Thus, the smallest molecules will enter all of the pores, while molecules of intermediate size, because of the velocity of the mobile phase, will not have sufficient time to diffuse into all of the pores into which they would fit normally and, therefore, will be retained less effectively. The result is a chromatogram that consists of an initial peak containing all of the totally excluded substances, followed by a group of peaks representing all of the substances that have been retained partially and separated, and finally another single peak caused by all of the totally included solutes.

The stationary phases used in this mode of partition chromatography are of two types.

The *soft gels* are made usually from cross-linked carbohydrates, such as dextran (Sephadex), agarose (Sepharose), or polyacrylamide (Bio-Gel), the use of which was described first by Porath and Flodin.[14] These are very hydrophilic, and before the column can be packed, they must be mixed with the mobile phase until they have *imbibed* enough liquid to become swollen completely. Once the column has been packed, the composition of the mobile phase cannot be altered, because this would change the amount of imbibed solvent, resulting in shrinking of the bed, or in further swelling that may burst the column. These gels are used with mobile phases that are primarily aqueous, and the technique is called *gel filtration.* Because of the low structural strength of the soft gels they cannot be used under high pressure. Size-exclusion media made from silica gel with controlled pore sizes have been developed for HPLC; they do not deform under pressure and can be used with aqueous or nonaqueous mobile phases.

The *semirigid* or *rigid gels* consist of materials such as cross-linked polystyrene, controlled-porosity glass beads, or alkylated
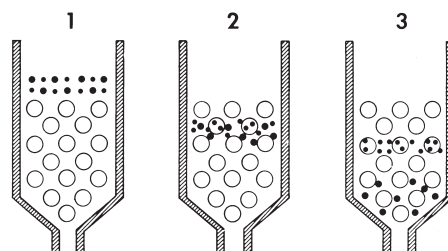
Figure 33-8. Size-exclusion chromatography. Small, soluble molecules (•) penetrate the pores of the gel (○) and are retarded. Macromolecules (●) are excluded from the gel matrix and elute first.

dextran. These can be used for the separation of organic-soluble polymers using nonaqueous mobile phases, such as chloroform, acetone, pyridine, or tetrahydrofuran. This technique is called *gel permeation,* and was described first in 1964 by Moore.[15]

Ideally, the only separation mechanism occurring in size-exclusion chromatography is that which depends on the diffusion of the solutes into and out of the pores. However, depending on the nature of the solute and the stationary phase, other retention mechanisms such as ion exchange, hydrophobic partitioning, or hydrogen-bonding may have an effect on certain solutes. These can result in long retention times, irreversible adsorption, or loss of activity in biological molecules. Such difficulties can be minimized by changing ionic strength or pH of the mobile phase to reduce charge effects, or by using additives such as detergents that modify the shape and charge of biological molecules.

Desalting is frequently necessary for the purification of biochemicals that have been separated from tissue using techniques involving buffers and precipitating reagents. In this procedure, a gel with a fairly low exclusion limit (ie, equivalent to a molecular weight of 1000–2000) is used. Because of the great differences in molecular weight between the biological molecules and the contaminating salts, short columns and high flow rates may be used. The macromolecules will be eluted in the void volume with little dilution, while the salts are retained on the column.

Concentration of dilute solutions of large molecules may be achieved with gels whose exclusion limit is less than the molecular weight of the substances involved. The solution is mixed with a small quantity of dry gel that will absorb 10 to 20 times its weight in water. Some salts and small molecules are taken up also, leaving the macromolecules in a solution of almost unchanged pH and ionic strength but significantly decreased volume.

Perhaps the greatest value of size-exclusion chromatography is for the fractionation and molecular-weight determination of macromolecules. It has been found that because the size of a molecule is approximately proportional to its molecular weight, M, the elution volume, V_E, can be expressed by

$$V_E = a + b \log M \tag{14}$$

where a and b are constants dependent on the mobile and stationary phases. To determine the molecular weight of a substance, the system must be calibrated by using an extremely large molecule, such as blue dextran, to establish the void volume of the system, and a substance such as deuterium oxide or sucrose to determine the retention time for a totally included solute. A series of standard proteins or polymers then is used to calibrate the region between these limits. A typical calibration curve of V_E versus log M for a series of protein standards is shown in Figure 33-9. Once the elution volume of the unknown compound is determined, the molecular weight can be estimated by interpolation.

Packing of the column is very critical in SEC. The column dimensions range from 600 mm × 16 mm to 1000 mm × 50 mm, and the volume of the loaded sample should not exceed 5% of the column volume for preparative runs and 1% for analytical applications. The recommended flow rates, which depend on the column diameter, should not be increased. In general, running the column at a low flow rate results in higher resolution, but diffusion may occur, when the flow rate is too low.[16]

Size-exclusion chromatography is used most often in procedures involving large biological molecules such as proteins, nucleic acids, and polysaccharides, which are not chromatographed well by other techniques. Among the procedures for which these gels are useful are desalting, concentration, molecular-weight determination, and fractionation.

AFFINITY CHROMATOGRAPHY

In situations where very specific separations are desired, *affinity chromatography,* a highly specialized form of adsorption chromatography, may be employed. This technique makes use
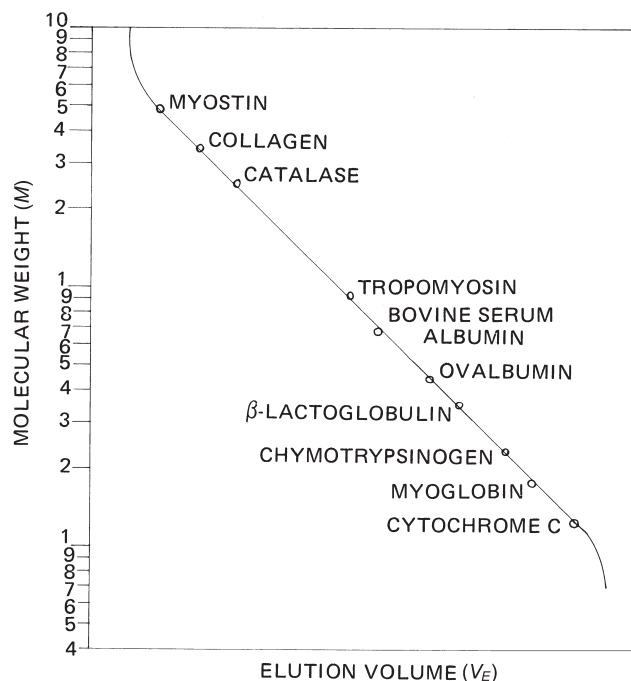
Figure 33-9. Typical size exclusion calibration curve of elution volume, V_E, versus log M for a series of protein standards.

of a specific ligand, which has been immobilized by being bound chemically to an insoluble matrix, to adsorb reversibly a single molecular species from a mixture of solutes. This method differs from other modes of chromatography already discussed in that, rather than attempting to separate a mixture of solutes for qualitative or quantitative analysis, it is concerned only with removing a single species from the mixture. It achieves its greatest utility as a highly specific purification technique for biological molecules.

Affinity chromatography owes its high degree of specificity to the nature of the binding forces between the ligand and the substance to be purified. Many biological molecules, as a result of their unique structure and conformation, form strong, noncovalent bonds to related compounds, as a drug would to a cellular receptor. The interactions between ligand and target molecule can be a result of electrostatic or hydrophobic interactions, Van der Waals' forces and/or hydrogen bonding. Examples of this are found in the association of enzymes with coenzymes, antigens with antibodies, lectins with carbohydrates, or polynucleotides with nucleic acids. If either member of the above-mentioned pairs is bonded permanently to a chromatographic matrix, it will be able to remove the other from solution without interacting significantly with any other solute in the mixture. Because the ligand-target molecule binding is reversible, a suitable mobile phase can be passed through the column to dissociate the pair and elute the purified substance.

The elution can be achieved by either specifically using a competitive ligand, or non-specifically, by changing the pH, ionic strength or polarity.[17] Some typical biological interactions, frequently used in affinity chromatography are (1) Enzyme - substrate analogue, inhibitor, cofactor; (2) Antibody - antigen, virus, cell; (3) Lectin - polysaccharide, glycoprotein, cell surface receptor, cell; (4) Nucleic acid - complementary base sequence, histones, nucleic acid polymerase, nucleic acid binding protein; (5) Hormone, vitamin - receptor, carrier protein; (6) Glutathione - glutathione-S-transferase or GST fusion proteins; 7) Metal ions - Poly (His) fusion proteins, native proteins with histidine, cysteine and/or tryptophan residues on their surfaces. Examples of the application of affinity chromatography are (1) Isolation of human immunoglobulins; (2) Purification of fusion proteins; (3) Proteins and peptides with exposed His, Cys, or Trp; (4) DNA binding proteins and coagulation factors.[17]

EXPERIMENTAL FACTORS AND INSTRUMENTATION

CLASSIC COLUMN CHROMATOGRAPHY

The experimental setup for performing this type of chromatography is relatively simple. The column into which the stationary phase is packed consists of a glass or Teflon tube, typically 10 to 50 mm in diameter and 5 to 100 cm in length, although much longer and wider columns have been used for difficult separations and preparative work. The bottom of the column is fitted with a stopcock or another type of flow restrictor to provide control over the flow rate of the mobile phase. The packing material is supported inside the column by means of a fritted glass disk or a piece of glass wool.

The packing may be introduced into the column either as a dry powder or as a slurry suspended in the mobile phase. In either case, it is essential that the bed be formed evenly with no air bubbles or channels to disrupt the flow of the mobile phase. If it is packed dry, the stationary phase is introduced in small quantities and allowed to settle with the aid of gentle tapping or vibrating on the outside of the column. If a slurry-packing technique is used, the stopcock is left open to allow the solvent to flow through the solid material and solvent is added, as necessary, to prevent the column from going dry while it is tapped to dispel air bubbles. When the bed has reached the desired height, the stopcock is closed and a layer of mobile phase left at the top of the bed. In slurry packing, positive pressure, vacuum, or a tamping rod may be used to ensure that the material is packed firmly.

The sample is placed on the top of the column in either of two ways. It may be mixed with a small portion of the stationary phase that then is packed as before or it may be dissolved in the mobile phase and deposited on the packing after the mobile phase has been allowed to run a slight distance into the stationary phase. More mobile phase then is added and the stopcock is opened to allow flow to begin. The chromatogram then may be developed by allowing the solvent to flow from a reservoir under the force of gravity or by introducing it under low pressure using a peristaltic pump. The effluent from the column is collected in fractions as a function of time or volume, and the eluates tested for the presence of the various solutes.

HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

Because of the relatively high pressures necessary to perform this type of chromatography, a more elaborate experimental setup is required. Figure 33-10 shows the block diagram of a complete HPLC apparatus. All of these components are not necessary to achieve successful analyses (see caption).

The *solvent reservoirs* are glass or stainless-steel containers capable of holding up to 1–2 liter of mobile phase, which may consist of pure organic solvents or aqueous solutions of salts or buffers. The substances used to prepare these mixtures should be of the highest purity available because contaminants eventually will be deposited on the column and disrupt the chromatography. The mobile phases are filtered to remove particulate matter that may clog the system, and they also are degassed using vacuum, sonication, or sparging with helium to eliminate outgassing in the pump or detector.

Because the particles that are used to pack HPLC columns are small enough (<50 μm) to prevent solvent flow by gravity, pumps that develop pressures up to 5000 psi are needed to force the mobile phase though the column. Two types are available: mechanical, which deliver at a constant flow rate, and pneumatic, which produce a constant pressure. Of the mechanical pumps, the most frequently used is the reciprocating piston type in which a motor-driven cam drives a sapphire plunger into a small liquid-end chamber to force out the solvent. Check valves control the flow of solvent into and out of the liquid end and prevent backflow. Because the flow pulses every time the plunger moves in and out, the pressure variations may cause an unstable baseline; thus, these pumps usually are equipped with a pulse-damping device. They may have two liquid chambers arranged in such a manner that, while one is filling, the other is delivering.

The pneumatic pumps may be either the gas-displacement type, which uses direct pressure from a highly compressed gas to force solvent out of a tube, or the pneumatic-amplifier type in which compressed gas at a lower pressure impinges on the large end of a piston to force the smaller end to deliver the liquid. The amplification of the original gas pressure is proportional to the ratio of the areas of the two ends of the piston. The pneumatic pumps have the advantage of pulseless operation.

If gradient analysis is necessary to achieve a particular separation, the most common way of forming the gradient is to include a second reservoir and pump and a *gradient controller*. This is an electronic device that synchronizes the operation of the two pumps to provide a mobile-phase mixture of the desired concentration. For example, if a 50–50 mixture of the solvents in the two reservoirs is desired at an overall flow rate of 1.0 mL/min, the controller will adjust the rate of delivery of each pump to 0.5 mL/min. The individual solvents then are combined in the *mixing chamber* and delivered to the chromatograph. The controllers are able to provide linear, convex, concave, or step gradients, thereby yielding a solvent mixture of constantly increasing strength to enable the resolution of complex mixtures.

In contrast to this high-pressure method of forming gradients, a low-pressure technique also is used frequently by some instrument suppliers. In this method, the mixing chamber is situated before a single pump. As many as four reservoirs are
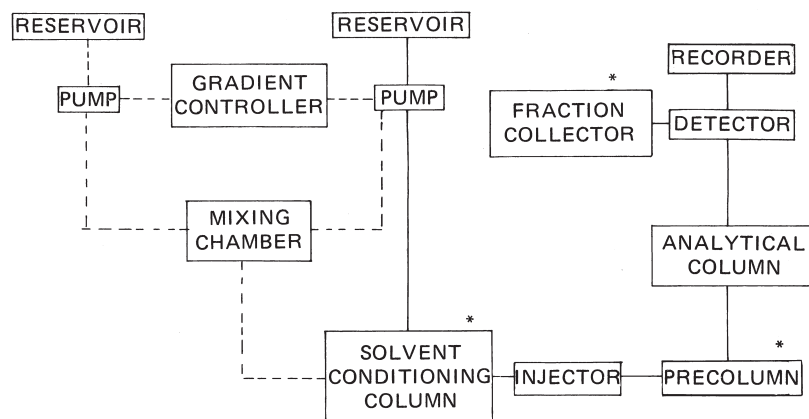
Figure 33-10. Block diagram of a complete HPLC. The items connected by dashed lines are necessary only for gradient elution. *These items are optional for both gradient and isocratic analysis.

connected to the mixing chamber, each reservoir having a remotely controlled valve. The gradient controller, either a separate device or a computer-resident software package, opens and closes the valves in the proper sequence so that the mixing chamber can generate mobile phase of the desired composition.

The next component, a *solvent-conditioning column,* is used only under special circumstances. Most HPLC column-packing materials are prepared from silica gel, which will dissolve slowly in solvents whose pH values are below 2 or above 7. This results in a shrinkage of the packing material, giving rise to void spaces in which separated solutes remix or are diluted, thereby leading to a loss of resolution. Therefore, to minimize this occurrence and to protect the expensive silica-based packing materials, a small column (5–10 cm) packed with HPLC-grade silica gel is inserted into the liquid stream after the pump but before the injector. The material in this column is dissolved preferentially, saturating the mobile phase and preserving the analytical column. Although there is some slight dissolution of the silica even in the pH range of 2 to 7, conditioning columns need not be used always and may be a disadvantage if fractions are collected with the object of recovering the solutes, as dissolved silica is difficult to remove from the solute.

The solute mixture is introduced into the chromatograph by means of a suitable *injection device.* Septum injectors are available, in which the sample solution is injected through a self-sealing rubber or Teflon disk using a microliter syringe. This may be done while the mobile phase is flowing or while it has been stopped temporarily. Although these devices are inexpensive and easy to use, it is difficult to achieve reproducible injections and automate their operation. Therefore, sample introduction is done mainly by using a rotary valve-and-loop injector. This consists of a stainless-steel and Teflon block that has been drilled to provide two alternate paths for solvent flow each selectable by a rotating valve. When the valve is in the "fill" position, the solvent flows through one path directly onto the column. In the other path there is a fixed-volume (10–1000 μL) loop of narrow-bore stainless-steel tubing, which is filled with the sample solution using a syringe or suction. When the valve is moved to the inject "position," the mobile phase path is diverted through the loop and washes its contents onto the column. The results obtained are very reproducible, and the injector can be automated by using a solenoid to change the valve position.

The next component in the instrument, called a *precolumn,* is optional and may be used for either of two reasons. When stationary phases consist of a thin layer of a liquid coated on a solid support, the liquid slowly dissolves in the mobile phase, causing a degradation of resolution. In this case, the precolumn will contain solid support coated with a higher percent of liquid phase than the analytical column to saturate the mobile phase and retard dissolution. Because most stationary phases used currently in HPLC are bonded permanently and not subject to dissolution, the precolumn is used mainly to protect the main column by trapping particulate matter and retaining substances, which would be irreversibly adsorbed on the analytical column. In this case it usually is called a *guard column.* The guard column is packed with a stationary phase identical to that in the main column, except that its particle size may be larger so that it will not restrict the flow. The larger material is relatively inexpensive and easy to pack so the contents of the guard column can be changed frequently. Because of its short length (2–10 cm), it usually does not affect the separation.

Most of the *analytical columns* used for HPLC are in straight lengths of stainless tubing, usually 5 to 25 cm in length, with an internal diameter of 2 to 4.6 mm, with highly polished interior walls. Compression end fittings attach the column to the HPLC apparatus. Stainless steel is useful with all organic and most aqueous buffers. However, chloride-containing buffers can slowly deteriorate the stainless steel, hence glass columns are recommended. Overall, stainless columns are used for most HPLC applications. Porous frits close the ends of column and retain the packing particles. Typically 2- and 0.5 μm- porosity stainless steel frits are used for 5- and 3-μm particles, respectively. Compression fitting column types are available with the wide selection of different packing materials. Well-made columns of this type provide highest level of performance and reproducibility.

The column packing particles for HPLC support utilize silica or porous polymers. Three types of particles are available for HPLC separations: (1) *Totally porous microspheres*–these are most commonly used because of the favorable properties in terms of efficacy, sample loading, durability, convenience, and availability. These particles are generally available in 5 μm size and all types of HPLC methods are developed with these materials. (2) *Micropellicular particles*–these have a solid core (silica or polymer based) with a thin outer skin of interactive stationary phase usually in 1.5 to 2.5 μm size. These particles are very useful for fast separations of macromolecules and generate very sharp peaks. (3) *Perfusion particles*–These contain very large pores eg, 0.4 to 0.8 μm in the particles. At high flow rates, the solutes can enter and leave this pore structure through a combination of convective and diffusion processes. These are mainly used in the preparative separation of macromolecules.[18]

Particle size of HPLC column packing play very important role in the effective separations. Particle diameters of about 5 μm represent a good compromise in analytical columns in terms of column efficacy, back pressure, and durability. Smaller porous particles are available for faster separations. Pellicular particles as small as 1.5 μm are useful for extremely rapid separation of macromolecules such as proteins. A narrow particle size distribution of < ±50% from mean in all materials ensure stable, high efficacy packed beds with minimum pressure drop. On balance, columns of the 3- or 5- μm totally porous microspheres meet the requirements of most HPLC separations.

It is possible to coat, to graft or even to encapsulate the chromatographic support with another material called stationary phase or, since bonding is such a common procedure, the bonded phase. Stationary phases with aliphatic hydrocarbon chains and phenyl moieties are used for reversed phase chromatography; amines and diols are used for normal phase chromatography; unmodified or alkylated amines are used for cation exchange chromatography; sulfonates or carboxylates are used for cation exchange chromatography; and affinity ligands, such as protein A and heparin, are used for affinity chromatography. Other ligands, including bovine serum albumin are used in chiral chromatography.

The most popular bonded phases are C1, C4, C8, and C18. Silica based adsorbents modified with trimethylchlorosilane (C1) and buthyldimethylchlorosilane (C4) have a few applications in HPLC, mainly for protein separation or purification. These adsorbents show significant polar interactions, although they do not have specific interactions caused by acidic silanols. Octyl (C8) and octadecyl (C18) modified adsorbents are the most popular ones. Almost 80% of all HPLC separations have been developed with these adsorbents. Propylphenylsilane ligands attached to the silica gel show weak dipole-induced dipole interactions with polar analytes. Usually this type of bonded phase is used for group separations of complex mixtures. Amino-compounds show some specific interactions with phenyl-modified adsorbent. A cyano-modified surface is very slightly polar. Columns with this phase are useful for fast separations of mixtures consisting of very different components. These mixtures might show very broad range of retention times on the columns. Amino-phase is a weak anion-exchanger. This type of column is mainly used in normal-phase mode, especially for selective retention of aromatic compounds. Diols are slightly polar adsorbent for normal-phase separations. These are useful for separation of complex mixtures of compounds with different polarity, and which usually shows a strong retention on unmodified silica.

The chromatographic column is often conceptualized as a stationary phase bed immersed in a rapidly flowing mobile phase, with stagnant pools of the mobile phase situated in the pores of the packing material.[18] Table 33-4 lists examples of some commonly available columns for different chromatographic modes and their base materials for various applications.[19,20]

Table 33-4. Examples of Some Commonly Available HPLC Columns

TYPE OF COLUMN	BASE MATERIAL	APPLICATIONS	SUPPLIERS OF COLUMN
Reversed-phase	Highly aqueous type Silica, C18 functional group Particle size mostly 5 μm	For analysis of polar compounds, columns can beused under 100% aqueous conditions	Astec, Waters, ShiseidoNacalai, Tesque, Thermo Electron Alltech, Jones Chromatography, GL Sciences
Reversed-phase	Regular type, Silica, C18 functional group Particle size 3–10 μm, mostly 5 μm	For retaining polar compounds in a wide range of applications	Agilent, Higgins Analytical, Cohesive Techlologies Siseido, Dionex, ESA, Waters, Interchim
Reversed-phase	Regular type, Silica, Perfluorinated phenyl functional group, Particle size 5 μm	Recommended for analysis of isomers, available in analytical, microbore, capillary, and preparative sizes	ES Industries
Reversed-phase	Monolithic polymeric silica column with endcapped C8 phase, C8 functional group	High-speed analysis with low back-pressure	Merck
Reversed-phase	Base deactivated type, Silica base material, C8 functional group, Particle size 5 μm	High-speed analysis, available in analytical, microbore, and reparative sizes	Macherey-Nagel
Reversed-phase	High stability, polyamide base material, Polyamide functional group, Particle size 5 μm	Hydrophilic polymeric bead with pH 0–14 stability; can withstand high pressures and temperatures; recommended for protein separations	Zordi
Reversed-phase	Regular type, Silica, C12 functional group Particle size 4, 10 μm	pH 1.5–10 stability; recommended for tryptic digests and small peptides	Phenomenex
	Polymer type, PS-DVB base material, PS-DVB functional group, Particle size 5 μm	For separating peptides, nucleotides, and other small molecules and also LC–MS.	Amersham Biosciences
Reversed-phase	Polar end capped, high aqueous type, silica base material, C18 functional group Particle size 2, 4, 10 μm	Polar end capping enables use in 100% aqueous mobile phase; recommended for retaining nonpolar and highly polar compounds and for LC–MS applications	Phenomenex
Reversed phase	Methacrylate/Methacrylate	Small protein and polypeptide analysis	Alltech Associates
Superficially Porous reversed phase	Silica/C3, C8, C18	High-speed protein analysis	Agilent Technologies
Normal Phase	Highly purified silica / 3μm to 300 μm particles	Nonpolar compound analysis, High sample recovery	YMC Inc., Perkin-Elmer
Normal Phase	Diol stationary phase, 6–30nm pore size and 5–50 μm particle size.	Diol columns also provide better reproducibility when compared with bare silica columns.	YMC Inc.
Normal Phase	μBondapak-NH$_2$, amine bonded silica	Nonpolar compound analysis	Waters
Normal Phase	Alkyl cyano phase on silica	Nonpolar compound analysis	BTR Seperations
Hydrophilic interaction	Silica/	Polar pharmaceutical analysis	Waters
Hydrophobic interaction	Silica/Diol	Protein analysis, desalting	Nacalai Tesque
Hydrophilic interaction	Silica/Sulfonic and quaternary amine	Peptide analysis	SeQuant
Ion exchange, Cation type	Polymethyl methacrylate, Sulfopropyl functional group, Particle size 7 μm	For separating proteins and peptides	Tosoh Biosep
Ion exchange, Cation type	Silica, Sulfonic acid functional group, Particle size 5 μm	For LC–MS analyses and ion-exchange separations of proteins, peptides	Thermo Hypersil-Keystone
Ion exchange, Anion type	PS-DVB base material, Cryptand functional group, Particle size 5 μm	Ror separating mono- and polyvalent anions in a single run and for determining trace anions in concentrated acids	Dionex
Ion exchange, Cation type	PS–DVB–methacrylate, Carboxymethyl or Sulfoethyl functional group, Particle size 5 μm	For separating biomolecules at most pH values and at high flow rates	OraChrom
Chiral	Silica/Amylose tris (3, 5-dimethylphenyl carbamate)	For use on compounds, with aromatic, amide, carbamate, ester groups, alkyl amines, and compounds with multiple stereogenic sites.	Chiral Technologies

(continues)

Table 33-4. *(continued)*

TYPE OF COLUMN	BASE MATERIAL	APPLICATIONS	SUPPLIERS OF COLUMN
Chiral	Silica/Cellulose tris (3, 5-dimethylphenyl carbamate)	Particularly effective for beta blockers, compounds with similar functionality and steroids Examples: alprenolol, atenolol, flavanone, metoprodol, oxprenolol, pindolol, propranolol.	Chiral Technologies
Chiral	Silica/Vancomycin aglycon	General chiral compound analysis	Astec
Chiral	Acidic chiral compound analysis	Silica / Quinidine or Quinoline	Bischoff Chromatography
Turbo flow columns	Polymer or silica	Turbulent flow chromatography	Cohesive Technologies
Polymeric	Polymethacrylate	Protein and peptide analysis	Supelco
Biomolecules	PS–DVB / PS–DVB	DNA fragment analysis	Biochrom Labs
Trapping column	Silica, DVB / Various	Biological sample analysis	Optimize Technologies
Environmental	Silica / Proprietary	Polyaromatic hydrocarbon analysis	Restek
Protein column	PS–DVB / PS–DVB	Proteomic analysis	Polymer Laboratories
Capillary column	Silica / C18	Proteomic analysis, drug discovery, and LC–MS	SGE
Proteomics	Silica / C4, C8, C18	Protein, peptide, and tryptic digest analysis	Micro-Tech Scientific
Trapping column	Strong cation exchange	Biological sample analysis	Scivex/Upchurch Scientific
Turbulent flow column	Silica/Sulfonic acid	Turbulent flow chromatography	Cohesive Technologies
Affinity column	PS–DVB–methacrylate / Protein A	Antibody purification	OraChrom
Affinity column	Porous glass / Protein A, protein G	Antibody purification	Millipore
Ion exchange	PS–DVB / Sulfonic acid	Amino acid analysis	Pickering Laboratories
LC–MS	Silica / C18, strong cation exchange	Protein and proteomic analysis	New Objective

Data from Majors RE. *LCGC North America* 2003; 21:240; and Majors RE. *LCGC North America* 2002; 20:248.

DETECTORS FOR HPLC

Devices used for detection of analytes in column chromatography typically rely on differences in the physical or chemical properties of the eluent and the analyte. Alternatively the solvent may be evaporated to allow detection by mass spectrometry, flame ionization or other detection modes.[21, 22] Table 33-5 lists the different types of detectors used in HPLC instrumentation.

Table 33-5. Detectors Used in HPLC Instrumentation

Optical Detectors
  UV/Visible
  IR/Raman
  Optical Activity
  Evaporative Light Scattering
  Refractive Index
  Luminescent
  Fluorescence/Phosphorescence
  Chemiluminescence/Bioluminescence
Mass Spectrometry Detectors
  Time-of-Flight/MALDI
  Fourier Transform Ion Cyclotron Resonance
  Mass Spectrometry (FTICR-MS or FT-MS)
  Electrospray/Thermospray
Elemental Detectors
  Atomic Absorption/Emission
  Inductively Coupled Plasma (ICP)-Mass Spectrometry
  Microwave-Induced Plasma
Electrochemical Detectors
  Direct current amperometry (DCA)
  Conductivity
  Coulometry
  Polarography

The most frequently used instrument is an ultraviolet-visible spectrometer that has been fitted with a flow cell of very small volume (8 μL). The simplest of these are fixed at one wavelength, usually 254 nm, because most aromatic organic compounds absorb strongly at, or near, this wavelength and the low-pressure mercury lamps used as light sources have a strong emission line at this point. Fixed wavelength models are also available at 280 nm, where the aromatic amino acids of proteins and peptides absorb, or at 214 nm, where isolated double bonds such as the carbonyl group absorb. The fixed wavelength detectors have the advantages of low cost and high sensitivity, being able to detect some compounds at the low-nanogram range. Sensitivity sometimes can be increased by using a variable wavelength detector, as it can be set to the exact point of maximum absorptivity for the solute.

More elaborate models, called *photodiode array detectors*, also are available. These can scan the entire UV spectrum repeatedly during the elution of a peak to determine if more than one substance is co-eluting.

A much more sensitive, but less broadly applicable, detector is the fluorescence spectrometer. Sensitivities in the picogram range can be attained with those compounds that fluoresce naturally or can be made to do so by derivatization. The less expensive models of these instruments are filter fluorimeters whereas the more sensitive ones use a prism or grating to provide monochromatic excitation and emission radiation. A benefit of fluorescence detector versus UV detector is its ability to discriminate analyte from interference or background peaks.

The most generally applicable detector available for use in HPLC is the differential refractometer, which is capable of measuring refractive index changes of 10^{-4} to 10^{-5} RI units. The other detector that is similar in function to differential refractometer is called evaporative light scattering detector (ELS). Each of these detectors has a similar sensitivity for typical samples, allowing the analysis of compounds present on the range of 0.1 μg/ml and higher. These are so called *universal de-*

tectors as that give response to all sample components. Universal detectors are used primarily in two applications: (1) for detection of very low level compounds and (2) to provide a more representative analysis of unknown samples by means of area normalization. Although these detectors react to almost all organic and inorganic compounds, these are not as sensitive as spectrometers. In addition, changes in ambient temperature cause severe drift and it cannot be used with gradient elution because in both cases differences in the RI are attributable to the solvent and not the solution.

Detectors based on electrochemical measurements such as amperometry, coulometry, polarography, or photoconductivity are used for readily oxidizable or reducible compounds such as the catecholamines. With the development of separator interfaces that remove part or all of the mobile phase and the concurrent use of narrow-bore ($\leq$2 mm) columns, spectrometric techniques such as mass spectrometry (LC-MS),and Fourier transform infrared (LC-FTIR) can be used as HPLC detectors. When a greater sensitivity is required than can be obtained from UV detection, the choice is usually fluorescence or an electrochemical detector. Electrochemical detection can be performed in either the oxidative or reductive mode, depending on the type of the analyte. Oxidative electrochemical detection is more commonly used because it is generally easier to (1) operate and run routine samples; (2) maintain the working electrode activity; and (3) avoid some of the preparative steps needed for routine reductive electrochemical detection. Reductive electrochemical methods also suffer from a poor signal / noise ratio due to reduction of dissolved oxygen in the solution.

DERIVATIZATION

Derivatization procedures are used with HPLC for a number of reasons:

- To allow chromatography of compounds that otherwise could not be detected by the instruments currently available, such as aliphatic amines, alcohols, and carboxylic acids.
- To improve resolution by adding a functional group that enhances the interaction of the solutes with the stationary phase, such as esterification of acids.
- To improve the sensitivity of the method, such as formation of fluorescent derivatives of amino acids.

Most of the derivatization reactions commonly used involve adding a substituted phenyl group to enhance detectability at 254 nm. These include the formation of *p*-bromophenacyl esters of alcohols, *p*-nitrobenzyl esters of carboxylic acids and *p*-nitrobenzyl oximes of carbonyls. Fluorescent derivatives (fluorescamine adducts of primary amines) are useful especially because they not only increase the sensitivity greatly, but they also allow selective detection of derivatizable compounds in the presence of coeluting substances that do not react with the reagent.

Derivatization may be done before the sample is introduced onto the column or after it has been eluted. Precolumn reactions provide a functional group, which may enhance the separation of the solutes as well as their detectability, for example, the formation of phthalaldehyde derivatives of amino acids.[23] Postcolumn derivatization allows the separation of the solutes based on their own functionalities but introduces a reagent into the column effluent before it reaches the detector in order to increase the sensitivity. Special items of equipment are available that have the capability of adding reagent, heating the reaction mixture, and providing a time delay to allow quantitative derivatization to occur before introducing the sample into the detector.

QUALITATIVE AND QUANTITATIVE ANALYSIS

The methods used for qualitative and quantitative analysis with HPLC are the same as those used with GC and the interested reader should consult that section for the relevant information.

RECENT DEVELOPMENTS IN HPLC

ION CHROMATOGRAPHY (IC)

Ion chromatography is a form of liquid chromatography that uses ion-exchange resins to separate atomic or molecular ions based on their interaction with the resin. This is a valuable technique for the analysis of inorganic and organic ions in trace amounts. Because there are other techniques, such as atomic absorption spectrometry, by which low levels of cations can be determined accurately, ion chromatography is most useful for the quantitation of anions. This method has a greatest utility for analysis of anions for which there are no other rapid analytical methods. It is also commonly used for cations and biochemical species such as amino acids and proteins.

The equipment usually consists of a resin or silica based low-capacity ion exchanger that permits the use of buffers of low ionic strength as eluants. This column is coupled to a conductivity detector. In those cases where the low signal generated by trace amounts of analytes is overwhelmed by the response due to the mobile-phase ions, a *suppressor column* may be used. This is a second ion-exchange column, which converts the ions of the mobile phase into molecular species of lower conductivity, thereby unmasking the analyte signal. The utility of ion chromatography has been demonstrated on samples as diverse as wastewater and biological fluids.

Ions in solution can be detected by measuring the conductivity of the solution. In ion chromatography, the mobile phase contains ions that create background conductivity, making it difficult to measure the conductivity due only to the analyte ions as they exit the column. This problem can be greatly reduced by selectively removing the mobile phase ions after the analytical column and before the detector. This is done by converting the mobile phase ions to a neutral form or removing them with an eluent suppressor, which consists of an ion-exchange column or membrane. For cation analysis, the mobile phase is often HCl or HNO_3, which can be neutralized by an eluent suppressor that supplies OH^-. The Cl^- or NO_3^- is either retained or removed by the suppressor column or membrane. The same principle holds good for anion analysis. The mobile phase is often NaOH or $NaHCO_3$, and the eluent suppressor supplies H^+ to neutralize the anion and retain or remove the Na^+.

SUPERCRITICAL FLUID CHROMATOGRAPHY

Supercritical fluid chromatography (SFC) is a relatively new chromatographic technique, having been commercially available in the recent years. As a result there is a large amount of research currently underway both in SFC method development and in hardware development. What differentiates SFC from other chromatographic techniques (GC and HPLC) is the use of a supercritical fluid as the mobile phase. This technique is based on the use as a mobile phase of a *supercritical fluid*, one held at or above its critical temperature, the point at which a gas cannot be liquefied no matter how high the pressure. The resulting liquid has density, viscosity, and diffusivity characteristics midway between its gaseous and liquid states. The lower density and viscosity, when compared to that of a liquid, allow faster separations than with ordinary HPLC, while the higher diffusion than that of a gas reduces longitudinal bandspreading.

The instrumentation employed is common to both HPLC and GC. Packed columns with reversed-phase bonded materials have been used as well as fused-silica open tubular capillary columns as long as 60 m. In addition to the usual HPLC detectors, the universal detectors used in GC such as the FID are also applicable. Solute retention may be influenced by gradient-programming of either the temperature or the pressure. The most commonly used mobile phase, carbon dioxide, has a critical temperature of 31° at 73 atmospheres and is an excellent solvent for many organic compounds as well as being inexpensive, nontoxic, and nonflammable.

Supercritical fluid chromatography has several main advantages over other conventional chromatographic techniques (GC and HPLC). Compared with HPLC, SFC provides rapid separations without the use of organic solvents. With the desire for environmentally conscious technology, the use of organic chemicals as used in HPLC could be reduced with the use of SFC. Because SFC generally uses carbon dioxide collected as a byproduct of other chemical reactions or is collected directly from the atmosphere, it contributes no new chemicals to the environment. In addition, SFC separations can be done faster than HPLC separations because the diffusion of solutes in supercritical fluids is about ten times greater than that in liquids (and about three times less than in gases). This result in a decrease in resistance to mass transfer in the column and allows for fast high resolution separations. Compared with GC, capillary SFC can provide high-resolution chromatography at much lower temperatures. This allows fast analysis of thermolabile compounds.

HYDROPHOBIC INTERACTION CHROMATOGRAPHY (HIC)

Separation and purification of active biological molecules such as proteins by normal reversed-phase techniques usually leads to denaturation and loss of activity due to the strong hydrophobic interactions with the supports or on account of the organic modifier (eg, acetonitrile or methanol) in the mobile phase. In HIC, supports are generally hydrophilic polymers, such as polyethylene glycol, onto which short hydrophobic ligands (eg, methyl, propyl, or butyl) have been bonded. Mobile phases are entirely aqueous with retention being controlled by the concentration of added salts such as ammonium sulfate or surfactants. High salt concentrations favor retention by increasing hydrophobic interactions with the column. Thus, the usual method of elution is gradient analysis with decreasing salt concentration. The weakness of the interaction with the stationary phase and the use of aqueous mobile phases enhance retention of biological activity.[24]

Hydrophobic interaction chromatography and reverse-phase chromatography (RPC) are closely related liquid chromatographic techniques. Both are based upon interactions between solvent-accessible non-polar groups (hydrophobic patches) on the surface of biomolecules and the hydrophobic ligands (alkyl or aryl groups) covalently attached to the gel matrix. The difference is that, adsorbents for RPC are more highly substituted with hydrophobic ligands than HIC adsorbents. The degree of substitution of HIC adsorbents is usually in the range of 10 to 50 mmoles/ml gel of C2–C8 alkyl or simple aryl ligands, compared with several hundred mmoles/ml gel of C4–C18 alkyl ligands usually used for RPC adsorbents. Consequently, protein binding to RPC adsorbents is usually very strong, which requires the use of non-polar solvents for their elution. RPC has found extensive applications in analytical and preparative separations of mainly peptides and low molecular weight proteins that are stable in aqueous-organic solvents. HIC is an alternative way of exploiting the hydrophobic properties of proteins, working in a more polar and less denaturing environment. Compared with RPC, the polarity of the complete system of HIC is increased by reducing the ligand density on the stationary phase and by adding salt to the mobile phase.[24]

Factors Affecting HIC

The main parameters to consider when selecting HIC media and optimizing separation processes on HIC media are:

1. Ligand type and degree of substitution: The type of immobilized ligand (alkyl or aryl) determines primarily the protein adsorption selectivity of the HIC adsorbent. In general, straight chain alkyl ligands show greater hydrophobic character than aryl ligands, which show mixed mode behavior where both aromatic and hydrophobic interactions are possible. It is also established that, at a constant degree of substitution, the protein binding capacities of HIC adsorbents increase with increased alkyl chain length. The choice between alkyl or aryl ligands is empirical and must be established by screening experiments for each individual separation problem. The protein binding capacities of HIC adsorbents increase with increased degree of substitution of immobilized ligand and reach a plateau at a certain high degree of ligand substitution but the strength of the interaction increases. Solutes bound under such circumstances are difficult to elute due to multi-point attachment.

2. Type of base matrix: The two most widely used types of support are strongly hydrophilic carbohydrates (eg, cross-linked agarose, or synthetic copolymer materials). The selectivity of a copolymer support will not be exactly the same as for an agarose based support substituted with the same type of ligand.

3. Type and concentration of salt: The addition of various salts to the equilibration buffer and sample solution promotes ligand-protein interactions in HIC by salting out process. As the concentration of such salts is increased, the amount of proteins binding also increases almost linearly up to a specific salt concentration and continues to increase in an exponential manner at still higher concentrations.

4. pH: The effect of pH in HIC is not straightforward. In general, an increase in pH weakens hydrophobic interactions, probably as a result of increased titration of charged groups, thereby leading to an increase in the hydrophilicity of the proteins. On the other hand, a decrease in pH results in an apparent increase in hydrophobic interactions. Thus, proteins, which do not bind to a HIC adsorbent at neutral pH, bind at acidic pH

5. Effect of temperature: The binding of proteins to HIC adsorbents is entropy driven, which implies that the interaction increases with an increase in temperature. The Van der waals attraction forces, which operate in hydrophobic interactions, also increase with increase in temperature. However, an opposite effect was reported some times indicating that the role of temperature in HIC is of a complex nature. This apparent discrepancy is probably due to the differential effects exerted by temperature on the conformational state of different proteins and their solubilities in aqueous solutions. In practical terms, one should thus be aware that a downstream purification process developed at room temperature might not be reproduced in the cold room, or *vice versa*.

6. Additives: Low concentrations of water-miscible alcohols, detergents and aqueous solutions of chaotropic ("salting-in") salts result in a weakening of the protein-ligand interactions in HIC leading to the desorption of the bound solutes. The non-polar parts of alcohols and detergents compete effectively with the bound proteins for the adsorption sites on the HIC media resulting in the displacement of the latter. Chaotropic salts affect the ordered structure of water and/or that of the bound proteins. Although additives can be used in the elution buffer to affect selectivity during desorption, there is a risk that proteins could be denatured or inactivated by exposure to high concentrations of such chemicals. However, additives can be very effective in cleaning up HIC columns that have strongly hydrophobic proteins bound to the gel medium.

CHIRAL CHROMATOGRAPHY

Many medicinally useful agents occur naturally as members of a racemic pair of chiral isomers. Frequently, the medicinal activity resides in one of the isomers, while the other has no appreciable activity or is toxic. To reduce the chance of untoward effects in patients, it is desirable to separate the mixture into its constituent isomers. The enantiomers can also differ in absorption, distribution, protein binding, and affinity to the receptor.[25] In order to separate the mixture (*racemate, which is a 1: 1 mixture of isomers*) into discrete isomers, it is necessary to react the sample with a chiral compound to form two *diastereomers* or *diastereomeric* complexes. Diastereomers have different chemical and physical properties and can be resolved to individual isomers. The separation process involves the use of GC, supercritical fluid chromatography and HPLC, but HPLC is most widely used of these methods.

Chiral seperation techniques can be classified as *indirect* and *direct* methods.[26] In the *indirect* separation technique, *chiral derivatizing agent* (CDA) is added to the mobile phase, which reacts with the isomers to form a mixture of two *diastereomers*. These have an additional chiral center, and their altered geometric structure permits separation (as they differ in physico-chemical properties) by conventional normal phase

or reversed phase HPLC. These loosely bound complexes are adsorbed to the stationary phase with different affinities and thus are separated. This approach circumvents the need for expensive columns with chiral stationary phases and is more flexible. This technique needs high enantiomeric purity and stability of the CDA. Another disadvantage is that isolation of the pure isomers requires removal of the chiral reagent. Examples of CDAs include (1) 1-(9-fluorenyl)-ethylchloroformate and o-phtaldialdehyde in combination with chiral thiols; (2) (O,O0-R,R)-diacylated tartaric acid anhydrides; (3) (1R,2R)- or (1S,2S)-N-[(2-isothiocyanato)cyclohexyl]-3,5 dinitrobenzoylamide (DDITC); (4) 1-(6-Methoxy-2-naphthyl)ethyl isothiocyanate (NAP-IT) and 2-(6-methoxy-2-naphthyl)-1-propylchloroformate (NAP-C).[27]

In the *Direct method,* a chiral-separating reagent can be bonded (covalently or ionically or physically coated) to the stationary phase to form a *chiral stationary phase* (CSP). Silica or aminopropyl silica is commonly used as a starting support material, but particles of a polymeric chiral stationary phase are also available. The chiral isomers associate loosely with the bound reagent to form dissociable diastereomers with different retention times. Resolution of chiral isomers relies on the formation of transient diastereoisomers on the surface of the column packing. The compound that forms the most stable diastereoisomer will be most retained, whereas the opposite enantiomer will form a less stable diastereoisomer and will elute first. Majority of chiral separations reported in the literature are direct separations. Several hundreds of CSP columns are commercially available, although many of these columns are similar in structure and enantioselectivity.[28] Table 33-6 lists different types of CSP columns according to their chemistry and mechanism of chemical recognition. Some columns are better able to separate wide range of sample types. A very rough order of CSP universality is, protein type > carbohydrate type > Pirkle type > cyclodextrin type.[18] Some suggestions for the choice of CSP for different chemical categories are listed in Table 33-7.

MICROBORE CHROMATOGRAPHY

The microbore chromatography procedure, also known as microscale HPLC or *capillary HPLC*, combines extremely high efficiencies with speed and economy of operation. The columns used in this technique are narrow-bore (1 mm or less) fused-silica or glass-lined stainless steel. The tubes may be coated with liquid phase or packed (densely or loosely) with microparticulate stationary phases of the same types used in ordinary HPLC. Special equipment is necessary for operation in this mode, including pumps that can deliver accurate volumes at flow rates of 50 to 200 μL/min, injectors that introduce samples of less than 1 μL, and detector flow cells with volumes of as low as 45 μL.

Transferring standard HPLC methods to Microbore HPLC allows huge increase in sensitivity. Sensitivity of most HPLC detectors is concentration dependent. Thus, the sensitivity of standard HPLC column can be increased by a factor of 20 when 4mm i.d. columns (typical flow rate: 1ml/min) are replaced by 1mm i.d. columns (typical flow rate: 50μl/min). In addition, columns can be connected in series to achieve additive increases in efficiency, a procedure that is not possible with ordinary HPLC columns. Other advantages include significant savings in solvents, as flow rates lower than 50 μL/min are required, low operational costs and finally the possibility of direct coupling to a mass spectrometer (LC-MS). The major disadvantage is a result of the low solute capacity of the columns, which makes preparative work impractical.

Microscale HPLC columns can be hyphenated to infrared (IR) spectrometer or mass (MS) spectrometer for compound detection and charecterization. Usage of microscale columns increases the mass sensitivity of a sample, IR and MS spectrometric techniques are hyphenated with microscale columns with great success.[31]

Increased mass sensitivity of micro-HPLC is significant in the analysis of biological samples. Microscale HPLC technique is successfully applied in the determination of bile acids in serum (detection levels, 0.13–0.28 pmole),[32] determination of dibutyl phthalate (DBP) in water (detection levels of tap water and commercially available purified water were 4.5 and 5.4 parts per billion, respectively),[33] determination of anti-oxidants in gasoline,[34] and determination of theophylline in serum and so on.

THE HIGH PERFORMANCE LIQUID CHROMATOGRAPHY METHOD DEVELOPMENT

High performance liquid chromatography is a widely used technique in the pharmaceutical and biopharmaceutical sectors. The HPLC method development essentially follows series of

Table 33-6. Different Types of CSP Columns for Chiral Chromatography

CSP TYPE	APPLICATION	MECHANISM OF CHIRAL SEPARATION
Pirkle type*		
	Beta blockers, Warfarin, Ibuprofen, and aryl-amides, aryl-epoxides aryl-sulphoxides	Hydrogen bonding, π-π interactions, dipole stacking
Cellulose/carbohydrate based	Small aliphatic and aromatic compounds, cyclopentenones, alkaloids, tropines, amines, beta blockers, beta lactams, dihydroxypryidines	Combination of attractive interactions and inclusion complexes
Protein based	Benzodiazepine, Warfarin and oxazepam, amines and acids, in general analytes with ionizable groups.	Combination of hydrophobic and polar interactions
Cyclodextrin	The selectivity of a Cyclodextrin Phase is dependent on the size of the analyte. Alpha-cyclodextrin will include single Phenyl groups or Napthyl groups end-on. Beta-cyclodextrin will accept Napthyl groups and heavily substituted phenyl groups. Gamma-Cyclodextrin is useful for bulky steriod-type molecules	Inclusion complexation and hydrogen bonding
Ligand exchange type	Mainly for separation of alpha- amino acids	Coordination complexes with copper or other metals
Macrocyclic antibiotics+	All amino acids, amino acid derivatives such as methyl esters and peptides	π-π interactions, hydrogen bonding, inclusion complexation, ionic interactions and peptide binding.

* The columns within this group are mainly the result of the work of Bill Pirkle
+ The antibiotics immobilized with silica Rifamycin, Vancomycin and Ticoplanin to form CSPs.

Table 33-7. Suggested Columns for Different Compounds Types

CLASS OF COMPOUND	TYPE OF CSP TO USE	COLUMN TRADE NAME
Acids	Proteins, cellulose/Amylose, Pirkle	OVM, AGP, BSA, Chiralcel OD-OJ, Whelk-O-1
Amino Acids	Crown ether, ligand exchange, cyclodextrins, protein	CrownPak CR (+), 1-hydroxyproline, BSA, Cyclobond II
Amines	Proteins, cellulose/amylose, Pirkle, cyclodextrins	OVM, AGP, BSA, Chiralcel OD-OJ, CTA, DNBPG, naphthyl alanine, Whelk-O-1, Cyclobond II
Alcohols	Proteins, cellulose/ amylose, Pirkle, cyclodextrins	OVM, AGP, BSA, Chiralcel OB, OD-OJ, CTA, DNBPG, naphthyl alanine, Whelk-O-1, β-Cyclobond II
Esters	Pirkle, cellulose	DNBPG, Whelk-O-1, CTA
Sulfoxides	Pirkle, cellulose, protein	DNBPG, Whelk-O-1, Chiralcel OA, OB, BSA
Carbamates	Pirkle	DNBPG, Whelk-O-1
Ureas	Pirkle	DNBPG, Whelk-O-1
Crown ethers	Cyclodextrins	β-Cyclobond II
Metallocenes	Cyclodextrins	β-Cyclobond II
Thiols	Pirkle	DNBPG, β-GEM-1
Amino Acids	Pirkle	DNBPG, β-GEM-1
Succinamides	Pirkle	DNBPG, β-GEM-1
Hydantoins	Pirkle	DNBPG, β-GEM-1
Binaphthols	Pirkle	DNBPG, β-GEM-1
β-Lactams	Pirkle, cellulose	DNBPG, Whelk-O-1, Chiralcel OC, OF
Succinamides	Pirkle	DNBPG, β-GEM-1
Polycyclic aromatic hydrocarbons	Cyclodextrins	Cyclobond Ac
Cyclic drugs	Protein	BSA, AGP, OVM
Aromatic drugs	Protein	BSA, AGP, OVM
Lactones	Cellulose	Chiralcel OA, OB
Cyclic ketones	Pirkle, cellulose	Whelk-O-1, Chiralcel OA, OB
Alkaloids	Cellulose	Chiralcel OC, OD, OF, OG
Dihydropyridines	Cellulose	Chiralcel OC, OD, OG, OJ
NSAIDS	Pirkle, cellulose, protein	Whelk-O-1, Chiralcel OJ, AGP
Oxazolindones	Pirkle	DNBPG

From Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*, 2nd ed. New York: John Wiley, 1997; 174. Copyright © 1997. Reprinted with permission of John Wiley & Sons, Inc.

steps as summarized in Figure 33-11.[29] In most cases, the desired separation can be achieved easily with only few experiments, and in other cases considerable number of experiments need to be performed. Before beginning a method development process for any given sample, the analyst needs to know several sample characteristics, eg, chemistry and physical properties of the sample, separation goal (whether qualitative or quantitative analysis), need for pretreatment of sample, detection method etc. The HPLC method development cannot be generalized, and there is no universal procedure that can be adopted for different types of samples. The development procedure varies according to the chemistry and nature of the analyte. The procedure given below gives more focus on analytes that follow reversed or normal phase separation.

Nature of the Sample and Defining Separation Goals

At the beginning of method development activity, information concerning the sample chemistry, composition, and properties has to be reviewed. These include sample solubility, number of compounds present in the sample, chemical structures, molecular weight of compounds, UV spectra, pKa of compounds and concentration range of compounds in the sample. Ideally, complete description of the sample is needed. For example, in a tablet containing the active and inactive ingredients, the goal of HPLC separation is primarily the assay of active drug. From the information on chemistry of the given sample, chromatographers can choose appropriate initial chromatographic conditions (ie, type of column and stationary phase needed, column dimensions, mobile phase composition, flow rate etc.). Many times the composition of many samples is not fully known at the beginning of HPLC method development (eg, samples containing impurities, degradation products, metabolites). In these cases, chromatographers follow an empirical procedure using a default column (mostly with reversed phase chromatographic conditions) and initial conditions as described under 'initial separation conditions'.

The chromatographer needs to define what is the goal of his analysis for the given sample . The chromatographic separation or analysis is carried out for one or more of the following reasons: (a) whether the goal is quantitative analysis of the active ingredient of the sample or qualitative detection of impurities in the given sample mixture, (b) number of samples to be analyzed at a given point of time (when a large number of samples must be processed at the same time, the run time can be decreased by decreasing column length and/or increasing the flow rate by compromising on the resolution of the samples), and (c) whether to resolve all the impurities or degradation products in a sample mixture; (many times it is essential to isolate all the impurities and degradation products from the active ingredient, but it may not be essential to resolve these impurities into individual components).

Sample Pretreatment and Need for Special Procedures

Sample preparation is an essential part of HPLC analysis intended to provide a homogeneous solution that is suitable for injection into the column. The sample should be prepared in such a way that it is relatively free from interferences and /or protect the column or equipment from damage. Furthermore, the sample solvent should be miscible in mobile phase without affecting the retention and resolution of the sample. Best results are often obtained when the composition of sample solvent is close to that of mobile phase since this minimizes base line upset and other problems. Some times it may be desirable to concentrate the analytes in the sample and / or derivatize them for improved detection or better separation.

Sample pretreatment includes large number of methodologies, as well as multiple operational steps and can therefore be a challenging part of HPLC method development. A sample pretreatment procedure should provide quantitative (>99%) recovery of analytes and involve minimum number of steps. The samples should be collected using a statistically validated process, stored in inert and tightly sealed containers. Several

sample pretreatment processes are available for making the sample suitable for HPLC analysis. One or more of these methods are employed in the sample pretreatment.

   (A) *Solid phase extraction:* Liquid is passed through solid phase, which selectively removes the analyte or interferences. Analyte can be eluted with a strong solvent and in some cases, interferences are retained and analytes are allowed to pass through solid phase. A wide variety of stationary phases is available for selective removal of desired inorganic, organic or biological analytes.

   (B) *Liquid-liquid extraction:* Sample is partitioned between two immiscible phases, which are chosen to maximize differences in solubility.

   (C) *Dilution:* Sample is diluted with a solvent compatible with HPLC mobile phase to avoid column overload or to be in linear range of detector.

   (D) *Evaporation:* Liquid is removed by gentle heating at atmospheric pressure with flowing air or inert gas or vacuum. Rotary vacuum evaporator is commonly used and automated systems (eg, Turbovap) are also available.

   (E) *Distillation:* Sample is heated to boiling point of solvent and volatile analytes are concentrated in vapor phase, condensed and collected.

   (F) *Microdialysis:* A semi-permeable membrane is placed between two aqueous liquid phases, and sample solute transfer from one liquid to the other progresses based on differences in concentration. Sample enrichment techniques such as solid phase extraction are required to concentrate the dialysate. Microdialysis is used for examination of extracellular chemicals in living plant or animal tissue using molecular weight cut off membranes. Molecular weight cut-off membranes are used on-line with micro-LC columns. These membranes some times are used on-line to deproteinate samples prior to HPLC since large proteins cannot pass through membranes

   (G) *Lyophilization:* Aqueous sample is frozen and water removed by sublimation under vacuum. For suspensions, filtration, centrifugation or sedimentation are carried out to remove particulate matter and the resultant solution is processed by one of the methods as described above. These are highly recommended to remove particulate matter and this prevents backpressure problems and preserves the column life.

Choosing Detector and Detector Settings

Before the first sample is injected during HPLC method development, we must be reasonably sure that the detection cell will sense all sample components of interest. In most cases, variable wavelength (spectrophotometric) or Photodiode-array (PDA) detectors are normally the choice because of their convenience and applicability for most samples. For many samples, good analytical results are obtained by careful selection of the detector wavelength. PDA detector permits simultaneous collection of chromatograms at different wavelengths during a single run, therefore providing more information on sample composition than is provided by a single wavelength run. The wavelength chosen for UV detection must provide acceptable absorbance by various analytes in the sample combined with acceptable light
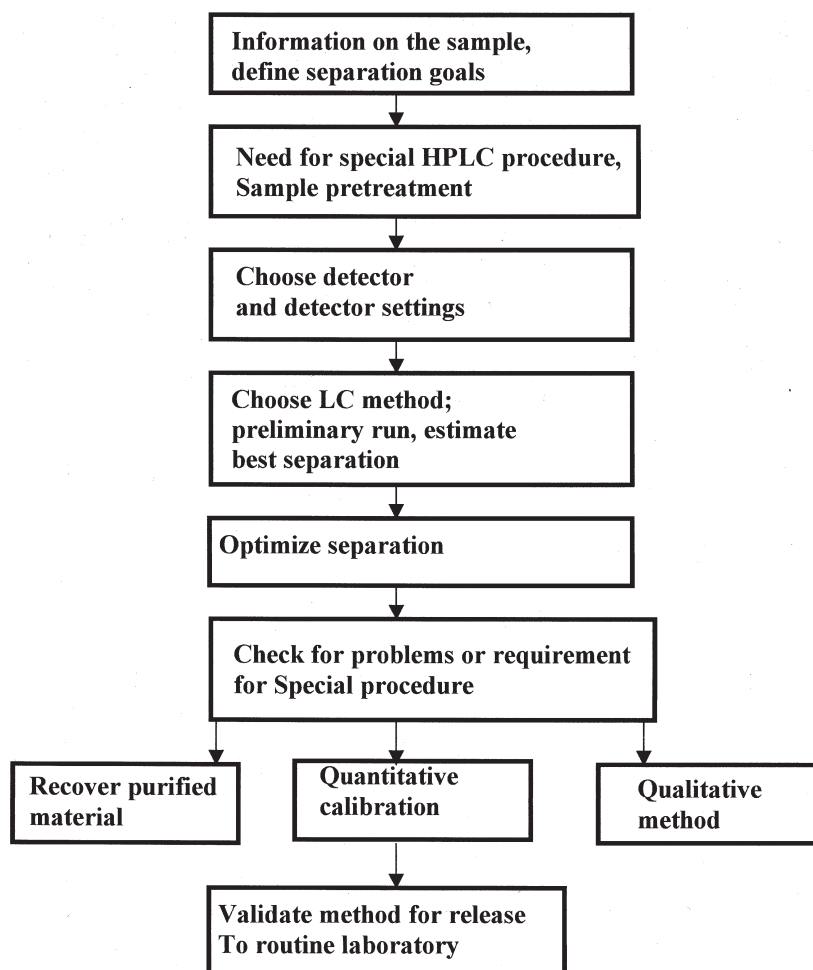
Figure 33-11. Steps in HPLC method development. From Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*, 2nd ed. New York: John Wiley, 1997; 174. Copyright © 1997. Reprinted with permission of John Wiley & Sons, Inc.

transmittance by mobile phase. For some samples, it is also important to select a wavelength at which sample interferences have minimum absorption. For UV detection to provide adequate sensitivity for the analysis of major sample components, molar absorptivity (ε) must be greater than 10. For trace analysis, values greater than 100 (preferably >1000) are usually required for UV detection. The organic compounds for which UV detection is completely unsuitable are saturated hydrocarbons and their amino and nitrile derivatives. Saturated hydrocarbons substituted by ether (-O-), hydroxy (-OH), chloro (-Cl), carboxy (-COOH), or ester (-COOR) groups have marginal absorptivity and may require detection at low UV values (210–185 nm). Since many mobile phases and degradation products in the sample absorb strongly in this range, analysis in this region is somewhat restricted. Compound types other than those mentioned above generally have larger absorptivity values and can be detected at higher wavelengths (> 210 nm).

When the sample shows little or no UV response or analyte concentrations are too low for UV detection, other detectors such as electrochemical or fluorescence detector can be considered. Further, sample can be derivatized for enhanced detection. The use of mass spectrometer for HPLC detection (LC-MS) is becoming commonplace. A mass spectrometer can facilitate HPLC method development and avoid common problems by (1) identifying the components of individual peaks in the chromatogram, (2) distinguishing the analytes of interest from interfering substances, and (3) recognizing unexpected and overlapping interference peaks to avoid a premature finish to method development.

Choice of HPLC Method and Choosing Column and Initial Separation Conditions

HPLC analytes can be classified as *regular* or *special* samples. Regular samples are defined as typical mixtures of small molecules <1000 Da that can be separated using more-or-less standardized starting conditions.[29] Regular samples are either neutral or ionic. Ionic substances include acids, bases, organic salts (ionized strong acids or bases) or sometimes, the sample composition is completely unknown (ie, the sample could contain acids or bases). Initial separation of regular samples is achieved by *reversed phase* chromatography under standard column dimensions, solvent system and run conditions (Standard conditions include: 15 cm column length and 0.46 cm i.d., 5 μm particle size, C_{18} stationary phase, buffered acetonitrile mobile phase, pH 3.0, 1.5–2.0 ml/min flow rate, 35–45 ºC column temperature and 25–50μl sample volume).[29] A gradient run is preferable to isocratic run to resolve the sample mixture. The initial gradient should be 5% to 100% acetonitrile in 60 min. This first gradient run can be used to decide (a) whether isocratic or gradient elution is recommended for analysis and b) if special reversed-phase conditions will be needed. Using these conditions, first exploratory run is carried out and then improved systematically. If the typical reversed-phase conditions provide insufficient sample retention, it suggests the use of either ion-pair or normal phase HPLC. In some cases, the sample may be strongly retained with 100% acetonitrile as mobile phase suggesting the use of non-aqueous reversed-phase chromatography or normal phase HPLC.

Special samples are usually separated with a different column and customized conditions. Examples of some special samples and their HPLC requirements are:

Inorganic ions isomers: Detection is primary problem; use ion chromatography.

Enantiomers: Require chiral conditions for separation as described in chiral chromatography.

Biological (proteins, peptides and oligonucleotides): Several factors make this kind *'special'*: These are molecular conformation, polar functionality and hydrophobicity. These are successfully resolved by ion pair or ion exchange chromatography.

Macromolecules: These may be separated by column packing with large pores (>>10-nm diameter) by size exclusion chromatography as described earlier.

The column is the heart of HPLC separation processes. The availability of stable HPLC column is essential in developing a rugged reproducible method. Commercial columns can differ widely among suppliers and even between supposedly identical columns from a single source. Such differences can have serious impact on developing the desired HPLC method. For selecting column with appropriate bonded phase, the first requirement is to know the chemistry of sample components. Molecular struc-

Table 33-8. Column Selection Chart

	PARTICLE SIZE (MICRONS)			COLUMN LENGTH (MM)			COLUMN INNER DIA (MM)			SURFACE AREA (M²/G)		PORE SIZE (Å)			CARBON LOAD (%)		
	3	5	10	30	150	300	2	4.6	22.5	200	300	60	100	300	30	10	20
What Do You Need? Default column#		*			*			*		*			*			*	
High efficiency	*																
High capacity										*	*						*
Low back pressure			*	*													
High resolution	*					*					*						*
High sample loadability								*		*	*						*
Capability to analyze samples with greater than 2000 molecular weights														*			
High stability																	
High sensitivity				*			*										
Fast analysis				*						*					*		
Low mobile-phase consumption				*			*										
Stability at pH extremes																	
Fast equilibration				*						*							

A default column is good for most applications.

Table 33-9. Effect of Column Dimensions and Particle Physical Characteristics on the Chromatographic Separations[30]

Column dimensions:

The column dimensions are the length and inner diameter of the packing bed.

Short: Short columns are 30–50 mm in length. They provide short run times and low back pressure.

Long: Long columns are 250–300 mm in length. They provide higher resolution and longer run times.

Narrow: Narrow-bore columns have inner diameters smaller than 2.1 mm. They provide higher detection sensitivity.

Wide: Wide-bore columns have 10–22 mm inner diameters. They enable the loading of large samples.

Particle shape:

Particles are either spherical or irregular.

Spherical particles have reduced backpressures and provide longer column life when used with viscous mobile phases such as 50:50 (v/v) methanol–water.

Irregular particles have higher surface areas and higher carbon loads, and they generally produce higher capacity factors for potentially greater resolution.

Particle size:

The particle diameter range is 1.5–20 μm.

Smaller particles offer higher efficiency. Choose 1.5 or 3 μm particles for resolving complex, multicomponent samples. Otherwise, choose 5- or 10-μm packings.

Surface area:

The surface area is the sum of particle outer surface and interior pore surface in square meters per gram.

High surface areas generally provide greater retention, capacity, and resolution for separating complex, multi component samples.

Pore size:

The pore size is the average size of the particles' pores or cavities. They range in value from 60 Å to 10,000 Å.

Larger pores allow larger solute molecules to be retained through maximum exposure to the surface area of the particles. Choose a pore size of 150 Å or less for samples with molecular weights less than 2000. Choose a pore size of 300 Å or greater for samples with molecular weights greater than 2000.

Bonding type:

The bonding type is the attachment mode of each bonded-phase strand to the base silica.

Monomeric: Monomeric phases have single-point attachments of bonded phase molecules. Monomeric bonding provides faster equilibration and higher column efficiency.

Polymeric: Polymeric phases have multiple-point attachments of bonded phase molecules. Polymeric bonding offers increased column stability, particularly when used with highly aqueous mobile phases. Polymeric bonding also enables columns to accept higher sample loading.

Carbon load:

The carbon load is the amount of bonded phase attached to the base material, expressed as the percentage of carbon.

High carbon loads generally offer greater resolution and longer run times for hydrophobic samples.

Low carbon loads shorten run times and often show different selectivity.

Endcapping:

Endcapping is the capping of exposed silanols with short hydrocarbon chains after the primary bonding step. Endcapping reduces peak tailing of polar solutes that interact excessively with the otherwise-exposed silanols. Nonendcapped packings provide a different selectivity than that of endcapped packings, especially for polar samples.

Data from Young CS, Weigand R. *LCGC North America* 2002; 20:465.

---

tures for all sample components should be known and two molecular structures that are the most similar need to be identified. Analyte retention occurs as the functional groups in question interact with the stationary and mobile phases and promote resolution of the pair by a process of differential migration. The intermolecular forces involved are Van der Waals interaction, dipole–dipole interaction, hydrogen bonding and π-π interactions. *C18, C8, and C4* are non-polar phases and retention for these phases is based upon Van der Waals interactions with hydrophobic compounds. Because the C8 phase has approximately 40% to 50% of the carbon loading of a C18 phase, its hydrophobicity and hence it's hydrophobic resolving power is less than that of a C18 phase. *Phenyl* phases also are nonpolar and retention for these phases is a mixed mechanism of hydrophobic and π-π interactions. The overall hydrophobic retention of a phenyl phase is similar to that of a C8 bonded phase but unique selectivity rests in its π-orbital interaction with analyte electron-deficient functional groups. *Cyano* phases have intermediate polarity. Retention is a mixed mechanism of hydrophobic, dipole–dipole, and π-π interactions. These phases are best used for analyzing polar organic compounds and they are versatile enough for use in both normal and reversed-phase modes. The *amino* phase is a polar phase that can be used in both normal and ion exchange modes. Retention is caused by dipole–dipole interactions or acid–base interactions. Amino phases commonly are used for carbohydrate analysis but they also can be used for analyzing both organic and inorganic ions.

The effect of column format and particle physical characteristics on the chromatographic separation is shown in Table 33-8.[30] The chart uses a default column as a starting point (first row). This profile represents an average analytical column that is good for most applications: 150 mm × 4.6 mm, 5-μm particles, 100-Å pore size, 200-m^2/g surface area, 10% carbon load, monomeric bonding, and spherical particles. To use the chart effectively, match the method goals with individual particle

physical characteristics. Change only those physical parameters that are affected by specific method goals. For example, a goal of fast equilibration is best achieved by using a short 30 to 50 mm column with a silica surface area of 200 m^2/g. Assessing the most important method goals will lead to important decisions regarding the physical aspects of the column and conclude with the selection of an optimal column. Finally, recognize the optimum column as a possible compromise of method goals. For example, the optimum column for highest resolution of all sample components could sacrifice speed of analysis. This is because; resolution (Rs) is dependent in part upon the number of theoretical plates, which in turn is affected by column length and particle efficiencies. Table 33-9 describes the column and particle physical characteristics and how they affect chromatography. This gives the analyst information for deciding the column type and specifications of column that to be used for separation of given sample.

Optimizing Separation

The separation achieved in one or two runs usually will be less than adequate. After few additional trials, it may be tempting to accept a marginal separation, especially if no further improvement is observed. However experienced workers realize that a good separation requires more than minimal resolution of individual sample bands, particularly for a routine procedure used to analyze a number of samples. Specifically the experienced chromatographer will consider several aspects of separation summarized as follows[29]:

Resolution: Precise and rugged quantitative analysis requires that the base line resolution to be greater than 1.5.

Separation time: < 5 to 10 minutes is desirable for routine procedures.

Pressure: <2000 psi is desirable, <2500 psi is usually essential (new column assumed).

Peak height: Narrow peaks are desirable for large signal/noise ratios.

Solvent composition: Minimum mobile phase use per run is desirable.

The time required for separation of the given sample should be as short as possible. The run time goal should be compared with the 2-hour setup time typically required for HPLC procedure including mobile phase preparation, column installation, equilibration, base line achievement, replicate standards injected to confirm precision, reproducible retention, and acceptable separation. Thus, if only two or three samples are to be assayed at one time, a run time of 20 to 30 min is not excessive. When several samples are to be assayed, run times of 5 to 10 minutes is desirable.

Conditions for final HPLC method should be selected so that the operating pressure with a new column does not exceed 2500 psi and an upper pressure limit below 2000 psi is desirable. This is because during the life of a column, the backpressure may rise by a factor of as much as 2 due to gradual plugging of the column with particulate matter. Moreover, at a lower pressure, pumps, sample valves, and auto-samplers operate much better, seals last longer and columns tend to plug less thereby improving system reliability. For these reasons, a target pressure of less than 50% of the maximum capacity of the pump is desirable.

While changing the mobile phase, enough time must elapse for the column to come into equilibrium with the new mobile phase and temperature. Usually column equilibrium is achieved after passage of 10 to 20 column volumes of the new mobile phase through the column. However, this should be confirmed by carrying out a repeat experiment under similar conditions. When constant retention times are observed in two or three successive experiments, it can be assumed that column is equilibrated and the experiments are repeatable. Failure to ensure column equilibration and repeatable chromatograms can be a serious impediment to HPLC method development.

Check for Problems or Requirement for Special Procedure

Problems in the HPLC method development are mostly concerning with column and instrument usage and maintenance. Three most important kinds of problems in HPLC method development are: (1) variability in retention and resolution, (2) band tailing, and (3) short column lifetime.

Columns must maintain reproducible retention and resolution during use, otherwise the accuracy and precision of the method are compromised and new columns may be required frequently. Sometimes, a new column may give unsatisfactory separation. This means that the operating conditions must be modified to re-establish the required separation. The main reasons for variations in retention and resolution are: variation in support and bonding (from column-to-column), disturbance in the bed and loss of bonded phase (during column usage), changes in the instrument-to-instrument configuration and insufficient equilibration time between stationary and mobile phases. Problems associated with this kind of irreproducibility are usually solved by: (a) initially selecting a good column of less-acidic highly purified support (silica column) and maintaining the same stationary phase, particle size and column dimensions throughout the operation; (b) eliminating the chemical or silanol effects for silica based columns by using favorable mobile phase conditions (pH, buffer type and additives); (c) making sure that the column is properly equilibrated with the mobile phase; (d) using proper laboratory techniques that ensure day-to-day operation; (e) using retention mapping to provide corrective action when required; and (f) stocking columns, or establishing a continuing supply of the same column.

Tailing peaks cause inferior separations and reduced precision. Column plate numbers and band resolution are over estimated when tailing peaks are involved. Tailing peaks can also trail into a closely eluting following peak reducing the ability to quantitate each peak accurately. Peak asymmetry or band tailing can arise from several sources: Bad column, plugged frit, build-up of waste in column inlet, sample overload, wrong solvent for sample, extra-column effects, chemical or secondary retention (silanol) effects, inadequate buffering and contaminating heavy metals. During use, columns can develop severe band tailing usually due to void in the inlet of the column and/or a dirty partially plugged inlet frit. This can be eliminated by replacing the inlet frit of the column or reversing the direction of flow through the column. Purging the column with a strong solvent sometimes can eliminate the buildup in the column (eg, dichloromethane), methanol and ammonium hydroxide (96%, 4%, 0.1% respectively are often effective for reversed phase columns and methanol (100%) for normal phase columns).

Columns for normal phase chromatography are more stable (>1 year life, when used with clean samples) than are columns used for other HPLC procedures. Polymeric ion exchange columns display similar stability. On the other hand, silica based columns for reversed phase chromatography, ion pair and ion exchange chromatography are less rugged in the aqueous environments. Columns degrade for several reasons: (a) Partially blocked (plugged) frit or column bed; (b) adsorption of sample impurities; (c) initially poorly packed column; (d) mechanical or thermal shock creating voids; and (e) chemical attack on support or stationary phase. The increase in column backpressure, tailing bands, decrease in retention are some symptoms imminent to column 'death.'

Some steps for ensuring best column life-time and performance are to: (a) use well-packed columns; (b) minimize pressure surges to avoid mechanical and thermal shock; (c) use guard column and an in-line filter; (d) flush column frequently with strong solvent; (e) pretreat samples to minimize particulates; (f) use stable stationary phase (C_{18} is the best); (g) use organic buffers when operating at intermediate pH (pH 6.0–8.0); (h) use column temperatures of less than 40° C; (i) keep mobile phase pH between 3.0 and 8.0 for most silica based columns; (j) add 200 ppm sodium azide to aqueous mobile phases and buffers; and (k) for overnight and storage, purge out salt and buffers, leave the column in pure organic solvent (preferably acetonitrile).

Finally, quantitation and method validation are an important part of the method development process. The key components of a method validation study are: (a) accuracy and precision; (b) linearity; (c) range; (d) limit of detection and quantitation; (e) specificity; (f) ruggedness and robustness; (g) stability of samples and reagents; and (h) instruments and system suitability criteria. In addition, method documentation data from interlaboratory crossover studies and techniques for determining equivalent performance are to be studied.

THIN-LAYER CHROMATOGRAPHY

Thin-layer chromatography (TLC) is a method of analysis in which the stationary phase, a finely divided solid, is spread as a thin layer on a rigid supporting plate; and the mobile phase, a liquid, is allowed to migrate across the surface of the plate. It differs from the techniques previously discussed in that the separation does not take place in a closed column, but rather on a planar surface; and the mobile phase does not flow under the influence of gravity or high pressure, but is drawn across the plate by capillary action. Although separation efficiencies equivalent to those obtained with gas or high-pressure liquid chromatography cannot be obtained by this method, it has the advantages of speed, versatility, and simplicity.

A wide variety of stationary phases are available in size ranges suitable for use in TLC. Because the mechanism of this method is essentially the same as that of liquid-column chro-

Table 33-10. Stationary and Mobile Phases Used in TLC

TECHNIQUE	STATIONARY PHASES	MOBILE PHASES
Adsorption	Silica gel	Alumina
Charcoal	Nonpolar or polar organic solvents	
	Polyamide	Polar organics
Partition	Cellulose	
Silica gel	Mixed aqueous, organic solvents	
Reversed phase partition	ODS silica gel	
	Coated silica	
Acetylated cellulose	Mixed aqueous, polar solvents	
Ion exchange	Ion-exchange resins	
	DEAE-and CM-cellulose	
	Buffered aqueous solutions	
Size exclusion	Dextran gels	Aqueous buffers

matography, the only distinction being that the separation takes place on a flat surface, the same modes used in liquid chromatography—adsorption, partition, ion-exchange, and size-exclusion—are available for thin-layer separations. These processes are listed in Table 33-10, along with some of the more commonly used mobile and stationary phases.

Silica gel, the most frequently used stationary phase, is employed as such for adsorption TLC and modified for reversed-phase separations by coating with a thin layer of a nonpolar substance, such as silicone oil, or by binding a nonpolar functional group to it, such as octadecylsilyl (ODS). The surface of silica is acidic due to the presence of many silanol hydroxyl groups; therefore, it is best suited to the analysis of acidic compounds. It also is preferable for polar compounds such as amino acids and sugars. Alumina (aluminum oxide) has a basic surface and is chosen over silica gel for the separation of basic and weakly polar compounds.

Polyamide (nylon) is a long-chain polymer that, because it has many free amide and carboxyl groups on its surface, is an adsorbent with strong hydrogen-bonding abilities. It will readily bond phenols, carboxylic acids, quinones, and nitro compounds, all of which require polar solvents such as methanol and dimethylformamide to displace them. Less active and less frequently used sorbents are calcium phosphate, calcium carbonate, and diatomaceous earth. Cellulose, a polysaccharide, has numerous neutral hydroxyl groups on its surface and can adsorb water or polar solvents by hydrogen-bonding, making it useful for partition TLC.

To ensure that the stationary phase adheres firmly to the backing plate and does not flake off during the development, binders such as calcium sulfate (gypsum), starch, or carbomethylcellulose are added to the adsorbent.

The mobile phases in TLC are identical to those used in liquid chromatography, and can be chosen using the eluotropic series shown in Table 33-3. If possible, it is preferable to use a single solvent to develop the chromatogram rather than a multicomponent mixture, because solvents are adsorbed preferentially by the stationary phase, and as the mixture moves up the plate, the composition of the mobile phase is always changing. Compounds that travel a greater distance up the plate, therefore, will be exposed to a different mobile phase than those that are retained strongly. The solvents also should be volatile so they can be evaporated from the plate after the development is completed.

The selection of the optimum solvent or mixture for use as the mobile phase depends also on the nature of the solutes and stationary phase and largely is empirical. A useful procedure for initial trials is to run two separate plates, one using a very polar solvent (eg, ethanol), and the other employing a nonpolar

liquid (eg, hexane). After observing which type of mobile phase moves the solutes from the origin and determining their k' or R_f values, the solvent may be modified to increase selectivity and resolution in a number of ways. The polarity may be altered by adding other solvents chosen by consulting tables of strength or dielectric constant. Substances with functional groups similar to those of the solutes, such as ethers, alcohols, or carboxyls, may be added to increase the R_f value by promoting solubility in the mobile phase. Acids or bases (acetic acid or ammonia) may be added to affect the charges on the solutes to prevent tailing.

PREPARATION OF PLATES

Plates with dimensions of 20×20 cm are necessary to attain the greater efficiency required for more difficult separations. These are usually made of glass, but plastic, stainless steel, or aluminum backings also are used. The material must be cleaned scrupulously to prevent interaction of the solutes with contaminants on the backing.

To reduce band-broadening, the stationary phase should consist of small particles of uniform size so as to provide a large area for interaction and a small void volume. The particles are mixed with water or an organic solvent to form a slurry, a suitable binding agent is added, and fluorescent indicators such as zinc silicate may be included to aid in detection of the solutes after the development. The slurry is coated on the plates using a spreader, which will apply a uniform layer of adsorbent of desired thickness over the surface of entire plate and the plates are dried.

Instead of coating a plate with one sorbent, two different substances may be applied simultaneously so that the layer is made of a gradient mixture of both. For example, silica gel and alumina can be used to prepare a pH gradient across the width of the plates. This may yield separations that otherwise would be impossible.

The thickness of the layer of stationary phase is important to the success of the chromatography, as excessively thick layers allow the solutes to diffuse laterally and, as in liquid column chromatography, band-broadening results. Layers from 0.1 to 2.0 mm in depth are used most often, with thinner ones (250 µm) being most suitable for precise separations and thicker coatings for preparative work, due to their greater solute capacity.

SAMPLE APPLICATION AND DEVELOPMENT

After the plates have been dried and conditioned, if necessary, in a controlled humidity chamber, the samples, which may range from a few µg to mg dissolved in 10 to 1000 µL of a volatile solvent, are spotted usually with a capillary tube or a microliter syringe. Samples may be applied as spots or as thin streaks, but it is essential that all of the solvent be evaporated between repeated applications and the area of sample application be kept as small as possible, because the bands will broaden as they travel up the plate.

For ascending development of the thin-layer chromatogram, the plate is placed in a rectangular jar that contains developing solvent to a depth of about 0.5 cm. The atmosphere of the jar should be saturated completely with the mobile phase before development, a process usually performed by lining the jar with a piece of filter paper that has been wet with mobile phase. The plate then is removed from the tank, the mobile phase front is marked by scratching the surface, and the solvent is evaporated in an oven or, if the sample is heat labile, in the air. To increase resolution, the techniques of *multiple development* and *two-dimensional development* have been used.

In the *multiple development,* after the plate is dried, it is returned to the chamber and redeveloped in the same direction, using the same mobile phase. The process may be repeated as many times as is necessary to ensure effective separation.

In two-dimensional TLC, the sample is applied as a small spot in the lower left corner of the plate, about 2.5 cm from each edge. After the plate has been developed in the usual manner, it is dried, rotated 90° counterclockwise, and placed in another chamber with a different developing solvent. The separated spots produced by the first elution are now located at the origin of the second. This method is useful especially for complicated mixtures containing many components or groups of substances with different functionalities, because selectivity effects of the mobile phases can be exploited more efficiently using two solvents.

DETECTION METHODS

Once the chromatogram has been developed, the solute spots must be made visible in order to determine their R_f values. If the substances are highly colored (eg, dye pigments), there is no difficulty in visual detection. Most organic compounds do not absorb visible light and routinely ultraviolet (UV) light is used to examine the separation compounds to detect light emission. *Fluorescence quenching* is a particularly useful technique for detection of compounds that absorb at 254 nm.

The two most frequently employed nonspecific methods involve the use of iodine vapor and charring of organic compounds. Iodine associates with practically all organic compounds, especially with unsaturated or aromatic compounds forming charge-transfer complexes. In any case the solutes will become visible as brown spots. Charring is a very widely employed technique for the detection of carbon-containing compounds, because it is effective for almost all organic compounds. The process involves spraying the plate with sulfuric acid, usually as a 50% (v/v) mixture with methanol, and then heating it in an oven at 110° for 10 to 30 min. The organic compounds are destroyed by the acid and a dark deposit of carbon (charcoal) remains at the spot. Though this method is effective for most organic solutes, it is destructive and, hence, cannot be used if the compounds are to be removed from the plates.

The more specific methods of detection involve spraying the plates with reagents designed to react with specific functional groups to produce visible derivatives. These reactions may produce products of three types:

- Those that are detected directly in visible light (2,4-dinitrophenyl-hydrazones of carbonyls)
- Those that absorb UV and quench fluorescence (benzoate esters of alcohols)
- Those that fluoresce directly (phthalaldehyde derivatives of amino acids).

Some of the more common derivatizing reagents, and the classes of compounds with which they react, are shown in Table 33-11.

The incorporation of radioactive elements, such as ^{14}C or ^{3}H, into the solutes provides another convenient method of detection, because special instruments are available that will scan a TLC strip and produce a chart recording similar to that obtained in GC.

Table 33-11. Commonly Used Derivatizing Agents

COMPOUND CLASS	REAGENT	COLOR PRODUCED
General	Iodine vapor	Brown
General	Sulfuric acid (50%)	Black
Acids	Bromcresol green	Yellow
Aldehydes and ketones	2,4-Dinitrophenylhydrazine	Yellow-red
Amines and amino acids	Ninhydrin	Fluorescent
Alkaloids	Mercuric nitrate	Yellow to brown
Barbiturates	Diphenylcarbazone	Purple
Carbohydrates	Aniline phthalate	Gray-black
Lipids	Bromthymol blue	Light-green
Steroids	Antimony trichloride	Various

QUALITATIVE ANALYSIS

In thin-layer chromatography, qualitative correlations of unknown compounds with standards are accomplished primarily by comparing the R_f value, which is the distance from the origin to the point of maximum intensity in the spot divided by the total distance of solvent travel. This method is used with great success in monitoring drugs of abuse in the urine of addicts undergoing treatment.

QUANTITATIVE ANALYSIS

Quantitation may be performed either while the solute is still on the plate or after it has been removed. The solute can be isolated from the plate in a number of ways. The area of adsorbent containing the substance can be removed from the plate by scraping or by aspirating it into a Pasteur pipet. The compound then is eluted from the adsorbent using a suitable solvent, and the solid stationary phase is removed by centrifugation or filtration. The solute then may be identified or quantitated by the usual spectrometric or chromatographic methods.

In those cases in which the solute band cannot be seen except by chemical reaction, underivatized solute may be obtained by running portions of the same sample in adjacent lanes or a sample in one lane and a standard in the next. After development, the sample lane is masked with a piece of glass and the remainder of the plate is sprayed with developing reagent to determine the location of the desired spot. The other lane is uncovered and the adsorbent removed in the area adjacent to the visualized solute.

Since the bonding between the solutes and adsorbents is frequently quite strong, complete removal from the stationary phase often is not achieved. Therefore, quantitative analysis of the substance while it is still on the plate is more reliable. Manual methods, such as comparing the spot sizes and intensities between unknown and standard, using a template, or tracing the spot outline on paper and weighing it, have been used, but they are tedious and give high levels of variability.

An automated method called spectrodensitometry is much more convenient and is capable of yielding quantitation in the submicrogram range. In this method, the plate is placed on a movable stage that is driven by a motor so that the lane of Spectrodensitometric measurements may be made on substances that are colored or absorb UV, those that have been charred, those that quench fluorescence, and even on photographs or x-ray films. A more detailed discussion of the quantitative aspects of densitometry can be found in the paper by Touchstone.[35]

The applications of thin-layer chromatography include (a) detection of narcotic and stimulant drugs in a single sample using a combination of R_f value and the various colors produced by over-spraying with different reagents,[35] (b) identification and isolation of active chemicals from plants and crude extracts, (c) determination of glycerol in tobacco, (d) determination of aflatoxins in foodstuffs, (e) determination of selenium after derivatization with 2,3=diaminonaphthalene in water and serum, (f) determination of vitamin B$_1$ in Pharmaceutical Products, and (g) determination of essential oils in herbal drugs etc.

PAPER CHROMATOGRAPHY

Although successful paper-chromatographic separations of dyes, salts, and other substances had been reported as far back as the middle of the 19th century, the method was not used widely until 1944 when Consden and co-workers[36] rediscovered and developed it just as they had done for liquid-partition and GC. They not only optimized the experimental procedure but also developed the theory of the separation process and formulated equations to describe the factors influencing the technique. Their work led to an appreciation of the method and its subsequent widespread application.

The stationary phase consists of a sheet of filter paper; the tightly bound water is the actual stationary phase and as a mobile phase passes over the surface of the paper, the solutes distribute themselves between the bound layer of water and the mobile-phase solvent. Therefore, the mechanism that predominates is liquid–liquid or partition chromatography, although adsorption to the cellulose surface also may occur. Papers especially impregnated to permit ion-exchange and reversed-phase chromatography also are available.

STATIONARY PHASE

The paper used in this method is prepared especially from cotton fibers and highly purified so as to be about 99% alpha-cellulose, which consists of polymers of glucose with molecular weights above 50,000. The chains of cellulose are bound together by hydrogen bonds in two different types of cross-linking. About 6% of the weight of the cellulose consists of water molecules permanently bound to the sugar hydroxyl groups, while another 10% to 20%, depending on humidity, are held more loosely. Because of the potential variability of the water content of the paper, moisture must be controlled carefully in its manufacture, storage, and use to achieve reproducible results. Some of the more important chromatographic papers and their characteristics are shown in Table 33-12.

Another variable introduced in the manufacture of the paper concerns the orientation of the fibers in the direction of motion of the machines that form it. Because the mobile phase travels across the paper by capillary action, the physical orientation of the channels is important in determining the rate of movement and, as a result, the flow is greater in the direction of the fiber orientation (*grain*) and slower perpendicular to it. In addition, there is a distance effect resulting in a slower flow as the distance from the origin increases.

Modified cellulose papers with a higher carboxyl content or attached ion-exchange functional groups (diethylaminoethyl, DEAE; or carboxymethylcellulose, CM) are available for the separation of cations, amines, and amino acids. For hydrophobic substances, cellulose-ester papers or those impregnated with mineral oil or silicone oil are used with polar organic solvents. Glass-fiber paper (Whatman GF/A) has been used, the main advantage being that it is not affected by reagents that are too corrosive for cellulose.

MOBILE PHASE

The solvents used for paper-chromatographic analysis are similar to those employed in other forms of partition chromatography. However, because the surface of the paper binds solutes strongly, mobile phases tend to be more polar than those used in thin-layer chromatography. Mixtures of alcohols, such as butyl or isopropyl, and water commonly are employed with ammonia or acetic acid added to control the charge on the solutes and reduce tailing.

Many organic substances are insoluble in water but soluble in polar organic solvents. For these compounds, paper impregnated with 20% to 40% of formamide in ethanol is used. In most cases, chloroform (for hydrophilic substances), benzene (for substances of medium polarity), cyclohexane (for hydrophobic substances), or a mixture of these solvents is used as the mobile phase. The advantages of these solvents are good separating ability and relatively short developing times, ranging from 1 to 4 hr.

SAMPLE PREPARATION AND APPLICATION

Drugs frequently are applied to the paper in solution in volatile solvents, such as ethanol, acetone, or chloroform, in quantities of 0.1 to 1000 μg, depending on the sensitivity of the detection method and the purpose of the analysis. In the determination of pharmaceutical or biological materials in which test substances occur at low concentrations, an extraction step generally must be employed, because substances like proteins, lipids, and inorganic ions may have undesirable effects when present in large amounts and, therefore, must be removed before the sample is applied to the paper. To enhance separation and identification, it often is advantageous to chromatograph derivatives when the original compounds are volatile.

Samples are applied at an origin that is located approximately 7 to 9 cm from the upper edge of the paper for descending development, 3 to 5 cm from the lower edge in ascending development, and on a circle with a radius of 1 to 3 cm for radial development. The optimum size of the spot varies from 3 to 8 mm in diameter, and adjacent spots should be 2 to 3 cm apart. Samples are applied with capillary pipets or microliter syringes, using multiple applications for large sample volumes, and drying each spot between applications.

DEVELOPMENT OF THE CHROMATOGRAM

The development of a paper chromatogram takes place in a glass or glass-lined stainless steel chamber of a size commensurate with the dimensions of the paper. This may range from a test tube for a small strip to a large cabinet or jar able to contain papers almost 2 feet long. The chamber must be kept sealed and saturated with the mobile-phase solvents. If the mobile phase is a mixture (eg, butanol–water) the two reagents are saturated mutually by shaking in a separatory funnel; the layers are separated, and the butanol layer transferred to the mobile phase reservoir in the chamber. The aqueous layer then is poured into a second container and placed in the chamber; the chamber is sealed and the vapors of the two solvents are allowed to come to equilibrium. The paper is spotted and placed in the chamber, but not yet allowed to contact the mobile phase, and the cellulose is permitted to equilibrate with the vapors.

The chromatogram is developed by allowing the mobile phase to travel over the surface of the paper in one of a number of ways: *ascending, descending, radial, linear horizontal,* or *spiral.*

Once the solvent has reached a point near the end of the paper, the process is stopped by removing the sheet from the chamber and allowing the solvent to evaporate. The spots then are made visible by methods similar to those employed in TLC, with the exception of charring, which is not useful because of the cellulose paper.

Qualitative analysis also is accomplished in the same manner as in TLC. The R_f values are determined and compared with standards, as are the results of specific derivatization reactions. Areas of the paper containing the compound of interest may be cut out and treated with a solvent to elute the substances. In descending chromatography, spots may be eluted off

Table 33-12. Types and Properties of Common Chromatographic Papers

PAPER	THICKNESS (MM)	WATER ASCENT[a]	DEVELOPMENT TIME[b]	CHARACTERISTICS
Whatman				
No. 1	0.16	140–220	15–16	Standard paper
No. 3MM	0.31	140–180	11	Preparative
No. 4	0.19	70–100	9	Fast
No. 31ET	0.50	60–120	4	Very fast
No. 54	0.17	60–120	6	Washed, fast
Schleicher & Schuell				
2040a	0.18	90–140	7	Fast
2043b (MGI)	0.23	220–260	15	Standard paper
2045b (GI)	0.16	300–400	45	Slow
2071	0.67	274–290	23	Preparative

[a] Time in minutes for water to ascend 30 cm up the paper.
[b] In hours, for the system: L-butanol:acetic acid:water (4:1:5).

the paper and collected in small containers at the bottom of the chamber.

Quantitative analysis may be accomplished by comparing spot size and intensity with standards developed under identical conditions, by densitometry, or by subjecting the material to standard spectrometric methods after eluting it off the paper.

ELECTROPHORESIS

Electrophoresis is defined as the migration of charged molecules under the influence of an external electric field. Since its introduction in 1937 by Tiselius for the purification of proteins, it has been used widely, especially for the separation of complex mixtures of biological substances such as proteins, nucleic acids, and polysaccharides. The name *electrochromatography* has been used for this process because, in some cases, as in chromatography, a narrow zone of solute is applied to a support, and migration in the electric field is influenced by the adsorptive or steric exclusion properties of the support. Electrophoresis is discussed at this point because some of the techniques are similar to chromatographic techniques with which they are combined readily.

The migration of particles in an electrophoretic system depends on properties of the particles as well as the instrumental system. Based on Stokes' Law, the mobility of a particle, μ, may be calculated from

$$\mu = \frac{Q}{6\pi r n} \tag{15}$$

where Q is the charge on the particle in esu, μ is in cm^2/volt-sec, r is the particle radius in cm, and n is the viscosity of the medium in poises.

For ions and peptides with a molecular weight of at least 5000 that do not obey Stokes' Law, Equation 16 is valid:

$$\mu = \frac{Q}{A\pi r^2 n} \tag{16}$$

where A has a value that ranges from 4 to 6 and is related to the particle shape.

Solution conditions are important variables. The solution pH determines the nature of species. For example, an acidic pH would favor protonation of basic centers of a protein, resulting in a positively charged molecule, whereas an alkaline pH leads to loss of protons from the protein, producing a negatively charged molecule. It is not desirable to choose a pH such that the protein is at its isoelectric point and exists as the uncharged zwitterion, a species not mobile in the imposed electrical field.

Electrophoretic mobility decreases with the supporting electrolyte ionic strength. Generally, the ionic strengths employed in electrophoresis range from 0.01 to 0.10. The temperature of the solution is important because the solution viscosity varies with temperature and the mobility increases with temperature. Because heat is generated during the electrophoretic process, this must be provided for in apparatus design and in experimental conditions.

The phenomenon of electroendosmosis arises because the solution itself migrates in an electrical field. This migration, which results from surface charges on the apparatus walls, usually is increased when a gel is added to stabilize the electrolyte and prevent the mixing of separated zones because of thermal gradients or diffusion. The stabilizing media develop a negative charge that causes the electrolyte and all zones, even the neutral compounds, to be carried to the cathode. Electroendosmosis effects are large with agar gels but small with polyacrylamide gels.

When no stabilizing medium is present or when a very porous system is used, the separations of species is related to the charge-to-size ratios as is seen in Equation 15. If stabilizing media are present, interaction of the species undergoing separation with molecules of the media introduce another consideration into the process.

Electrophoresis commonly is performed using one of two techniques.

In *moving-boundary* or *free-boundary electrophoresis*, the apparatus consists of a U-shaped tube with provision for introducing the cathode and anode electrodes into each of the arms. The sample solution is introduced and each arm is filled carefully with a buffer solution. If the sample consists of compounds with different mobilities, their migration may be observed as several moving boundaries. This method yields information on isoelectric points and mobilities of the compounds but usually is not useful for the isolation of the components because complete separation rarely is achieved. Several problems are associated with the technique, including stabilization of ion boundaries, boundary anomalies, and the need for specialized equipment.

Zone electrophoresis makes use of a stabilizing medium to minimize the problems associated with free-boundary electrophoresis. Many types of stabilizing media are available including paper, starch gels or blocks, cellulose, and agar or polyacrylamide gels.

One of the simplest procedures in electrophoresis involves spotting a mixture of solutes in the middle of a paper strip, moistening the paper with some electrolyte, and placing it between two sheets of glass. The ends of the paper strip extending beyond the glass plates are immersed in beakers of the electrolyte. A potential of approximately 5 V/cm of paper length is placed on this system, from a direct-current source. Electrophoresis is allowed to continue for a period of several hours. Usually, sufficient movement occurs in that time to obtain good separations, but longer periods sometimes are required.

Many other supporting media have been used for electrophoretic separations. *Cellulose acetate* strips, which are used widely in clinical laboratories, produce excellent separations of 7 to 9 protein fractions in a few hours. This material is exceedingly fine and homogeneous, and little *tailing* is encountered due to negligible adsorption. It especially is useful for separating α_1-globulins from albumin and provides a good background for staining glycoproteins (see Chapter 32).

Electrophoresis in compact gels, which depends at least in part on size-exclusion effects to achieve separation, is used frequently for the separation of proteins and nucleic acids. Although starch gels have been used in this respect, agar and especially polyacrylamide gels are employed most often. The degree of cross-linking of the individual acrylamide polymer strands may be varied during the preparation to produce gels of different pore sizes. This allows the separation conditions to be varied according to the size of the solutes in the analysis mixture. The overall migration in these gels is a combination of movement under the influence of the electric field and size separation by the pores of the gel.

The most frequently used technique of *polyacrylamide gel electrophoresis* (*PAGE*) is the discontinuous buffer system developed by Laemmli[37]. In this procedure the sample is placed on a *stacking* gel with a low level of cross-linking and, therefore, a large pore size. During movement through this gel, the sample is concentrated into a narrow band and then deposited onto a *separating* gel that has a higher cross-linking and smaller pore size. The separation of the solutes occurs in this phase.

In a special modification of this technique used for the separation of proteins, a detergent, such as sodium dodecylsulfate (SDS), is introduced into the buffer. This interacts with the proteins to produce particles of consistent shape and uniform negative charge so that separation occurs according to size alone. This enables the simple determination of molecular weight because the migration distance is proportional to the logarithm of molecular weight, as in size-exclusion chromatography.

Various methods have been used for the detection of the sample bands on the *electrophoretograms.* They include reaction with specific reagents such as Comassie Blue for proteins or ethidium bromide for nucleic acids to form derivatives that are detectable spectrally, general reactions such as staining with silver, or autoradiography using included radioactive labels.

In a common method of detection known as *blotting,* the macromolecules either are transferred passively or electro-eluted onto a suitable medium, such as nitrocellulose or a nylon membrane, following the electrophoretic separation. The membrane then is processed to detect the individual solutes. For nucleic acid separations, the membrane is developed using nucleotide probes of complementary sequence. These are known as Northern (for DNA) or Southern (for RNA) blots. Transferred proteins are detected with antibody probes in the technique of Western blotting.

Enzymatic and immunological methods also have been used to detect proteins following electrophoresis in gels. Immuno-chemical methods add an additional dimension to protein identification. Following electrophoresis in an agar gel backed with a microscope slide, an antibody is placed into a trough cut parallel to the direction of electrophoresis. The antibody and electrophoretically separated antigens diffuse toward each other resulting in precipitin arcs where antigen–antibody complexes form. This technique has been referred to as *immunoelectrophoresis.*

Polyacrylamide gels also have been used successfully for the fractionation of DNA and RNA. The technique yields separations that are superior to those obtained by zone centrifugation through sucrose density gradients; thus, the time of analysis is reduced greatly. Larger columns of starch, cellulose, or silica gel are suitable for preparative work, yielding highly purified fractions in sufficient quantity for chemical analysis.

A modification of one electrophoretic technique, called *isoelectric focusing*, rapidly is becoming an important tool for the separation of ampholytes, especially proteins. All proteins have an isoelectric point, pI, which is the pH value when the molecule has no net charge. When electrophoresis is run in a solution buffered at a constant pH, proteins having a net charge will migrate toward the opposite electrode so long as the current flows. The use of a pH gradient across the supporting medium causes each protein to migrate to an area of specific pH. Proteins are focused at the point in the gradient where they carry no net charge—the pI of the protein equals the pH of the gradient—thus resulting in sharp, well-defined protein bands.

Whereas separation by isoelectric focusing depends on the existence of a pH gradient in the system, the technique of *isotachophoresis* depends on the development of a potential gradient. A leading electrolyte (eg, chloride) with a higher mobility than the analytes, and a trailing electrolyte (eg, glycinate) with a lower mobility are used. The analytes are positioned between the electrolytes and, when the voltage is applied, they migrate in order of decreasing mobility. This establishes the potential gradient; from that point on, all the analytes move at the same speed. Isotachophoresis has been used for the separation of proteins as well as inorganic substances.

A technique that shares the attributes of both chromatography and electrophoresis is called *capillary electrophoresis* (CE). In this method, separation based on electrophoretic mobility takes place inside a capillary similar to those used in GC. The effective length of the capillary from the point of injection to the detector is commonly 25 to 50 cm, and the supporting electrolyte or "mobile phase" is usually a buffer, although a gel such as polyacrylamide may be used.

The apparatus used in this technique is very simple. The ends of the capillary are placed in buffer reservoirs, and these are established as anode and cathode by means of a DC power supply capable of delivering up to 30,000 V. At some point near the cathodic end of the capillary, a detector, usually an UV-visible spectrophotometer of the type used in HPLC, is placed so that a section of the capillary serves as its flow cell.

The sample is introduced at the anodic end, either by electromigration or positive pressure, and when potential is applied, net migration occurs in the direction of the cathode. Even substances with a net negative charge migrate in the direction of the cathode because of a phenomenon called the *electro-osmotic effect.* Because the capillary is made of silica, the surface contains many weakly acidic silanol groups. These dissociate in the presence of the buffer, leaving a negative charge at the surface and hydrated positive ions (H^+) in solution. When a potential is applied, the contents inside the capillary move toward the cathode, carrying along with them all of the analytes.

Neutral molecules move at the same speed as the electro-osmotic flow, while positively charged species move faster, their net speed being the sum of the electro-osmotic flow and their intrinsic electrophoretic mobility. Negatively charged molecules still move toward the cathode under the influence of the electro-osmotic flow, but they lag behind the other species. Within a group of similarly charged ions, separation is by electrophoresis.

The same modes as used in ordinary electrophoresis—zone, gel, isoelectric focusing, and isotachophoresis—are used in CE. However, none of these is successful in separating neutral molecules, a class exemplified by many pharmaceuticals. Although these substances will migrate under the influence of the electro-osmotic flow, they travel as a group and do not separate. Therefore, a mode called *micellar electrokinetic capillary chromatography* (MECC) is used in which a detergent, such as sodium dodecylsulfate (SDS), at a concentration above the critical micelle concentration, is included in the running buffer. As the resulting anionic micelles travel through the capillary, the neutral molecules partition in and out of the micelles selectively and separation is achieved.

Because the separation combines electrophoresis and chromatography, CE can achieve outstanding efficiencies even approaching 10^6 plates/m. However, sensitivity is lower than with chromatographic methods due to detection difficulties. The only detector in widespread use is the UV-visible spectrophotometer; however, because the optical path length across the capillary is on the order of 50 µm instead of the usual 1 cm, sensitivity to a particular compound is lowered by a factor of 200. However, research into improved systems, especially CE-MS, is actively continuing and applications in pharmaceutical analysis are increasing.[38]

REFERENCES

1. Tswett M. *Berichte der Deutschen botanischen Gesellschaft* 1906; 24:316.
2. *United States Pharmacopeia-National Formulary (USP26 NF21)*, United States Pharmacopeial Convention; Rockville, MD, 2003: 2126.
3. Martin AJP, Synge RLM. *Biochem J* 1941; 35:1358.
4. Martin AJP, James AT. *Biochem J* 1952; 50:679.
5. McNair HM, Miller JM. *Basic Gas Chromatography.* New York: John Wiley, 1997.
6. Berezkin VG, Viktorova EN. *J Chromatogr A* 2003; 985:3.
7. Eiceman GA, Gardea-Torresdey J, Overton E, et al. *Anal Chem* 2002; 74:2771.
8. Gloor R, Johnson E. *J Chromatogr Sci* 1977; 15:413.
9. Marchal L, Legrand J, Foucault A. *Chem Res* 2003; 3:133.
10. van Buel MJ, van der Wielen LAM, Luyben KCAM. In: Foucault AP, ed. *Centrifugal Partition Chromatography, Chromatographic Science Series.* New York: Marcel Dekker,1994; 68.
11. Foucault AP. *J Chromatogr A* 2001; 906:365.
12. Den Hollander JL. *J Chromatogr B* 1998; 711:223.
13. Sluyterman LA, Elgersma O. *J Chromatogr Sci* 1978; 17:150.
14. Porath J, Flodin P. *Nature* 1959; 183:1657.
15. Moore JC. *J Polymer Sci (Gen Pap)* 1964; 2:835.
16. Barth HG, Boyes BE, Jackson C. *Anal Chem* 1994; 66:595R.
17. Amersham Biosciences Corp., *Affinity Chromatography, Principles and Methods,* Piscataway NJ: Amersham Biosciences Corp, 2002; 7.
18. Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*, 2nd ed. New York: John Wiley, 1997; 174.
19. Majors RE. *LCGC North America* 2003; 21:240.
20. Majors RE. *LCGC North America* 2002; 20:248.
21. Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*, 2nd ed. New York: John Wiley, 1997; 59.
22. LaCourse WR. *Anal Chem* 2002; 74:2813.
23. Jones BN, Paabo S, Stein S. *J Liq Chromatogr* 1981; 4:565.
24. Builder SE, Amersham Biosciences Corp. *Hydrophobic Interaction Chromatography: Principles and Methods.* Piscataway NJ: Amersham Biosciences Corp, 1993; 13.
25. Eichelbaum M, Gross AS. *Adv Drug Res* 1996; 28:1.
26. Haginaka J. *J Pharm Biomed Anal* 2002; 27:357.

27. Szymura-Oleksiak J, Bojarski J, Aboul-Enein HY. *Chirality* 2002; 14:417.
28. Schurig V. *J Chromatogr A* 2002; 965:315.
29. Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*, 2nd ed. New York: John Wiley, 1997; 1.
30. Young CS, Weigand R. *LCGC North America* 2002; 20:465.
31. Jinno K, Tsuge S. In: Ishii D, ed. *Indroduction to Microscale High Performance Liquid Chromatography*. New York: VCH Publishers, 1988; 95.
32. Takeuchi T, Ishii D. *J High Resolut Chromatogr/Chromatogr Commun* 1983; 6:571.
33. Takeuchi T, Ishii D *J Chromatogr*. 1982; 253:41.
34. Ishi, D, Goto M, Takeuchi T. *J Chromatogr*. 1984; 291:398.
35. Touchstone JC, Levin SS, Murawec T. *Anal Chem* 1971; 43:858.
36. Consden R, Gordon AH, Martin AJP. *Biochem J* 1944; 38:224.
37. Laemmli UK. *Nature* 1970; 227:680.
38. Rabel SR, Stobaugh JF. *Pharm Res* 1993; 10:171.

BIBLIOGRAPHY

Cazes J, Scott RPW. *Chromatography theory*. New York: Marcel Dekker, 2002.

Deyl Z, ed. *Electrophoresis—A Survey of Techniques and Applications*, Vol 18. *Journal of Chromatography* Library. New York: Elsevier, 1979.

Dilts RV: *Analytical Chemistry*. New York: Van Nostrand, 1974.

Frei RW, Lawrence JF. *Chemical Derivatization in Analytical Chemistry*, vol 1, *Chromatography*. New York: Plenum, 1981.

Fries B, Sherma J. *Thin Layer Chromatography, Techniques and Applications*, 3rd ed. New York: Dekker, 1994.

Giddings JC. *Advances in Chromatography*. New York: Dekker, continuing series starting in 1965.

Giddings JC. *Dynamics of Chromatography*, Part 1. New York: Dekker, 1965.

Grob RL. *Modern Practice of Gas Chromatography*, 3rd ed. New York: Wiley, 1995.

Heftmann E. *Chromatography: Fundamentals and Applications of Chromatographic and Electrophoretic Method*. New York: Elsevier Science, 1983.

Satinder A. *Chromatography and Separation Science*. Boston: Academic Press, 2003.

Cserhati T. Esther F. *Chromatography in Environmental Protection*. Australia: Harwood Academic Publishers, 2001.

Cserhati T, Esther F. *Chromatography in Food Science and Technology*. Lancaster, PA: Technomic, 1999.

Kuksis A. *Chromatography of Lipids in Biomedical Research and Clinical Diagnosis*. New York: Elsevier Science, 1987.

Satinder A. *Chromatography of Pharmaceuticals: Natural, Synthetic, and Recombinant Product*. Washington: American Chemical Society, 1992.

Miller JM. Separation Methods in Chemical Analysis. New York: Wiley, 1975.

Reed E, ed. *Assay of Drugs and Other Trace Compounds in Biological Fluids,* vol 5, *Methodological Developments in Biochemistry*. Amsterdam: Elsevier, 1976.

Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC Method Development*. London: Wiley, 1997.

Snyder LR, Kirkland JJ. *Introduction to Modern Liquid Chromatography*, 2nd ed. New York: Wiley, 1979.

Snyder LR. *Principles of Adsorption Chromatography*, vol 3 of Giddings JC, Keller RA, eds. *Chromatographic Science Series*. New York: Dekker, 1968.

Touchstone J, Sherma J. *Techniques and Applications of Thin Layer Chromatography*. New York: Wiley, 1985.

Touchstone JC, Dobbins MF. *Practice of Thin Layer Chromatography*, 2nd ed. New York: Wiley, 1983.

Zweig G, Sherma J. *CRC Handbook of Chromatography*. Cleveland, OH: CRC Press, 1972.

Instrumental Methods of Analysis

Ajai K Chaudhary, M Pharm, PhD

Gurkeerat Singh, M Pharm, PhD

Gregory A Stephenson, PhD

Bradley L Ackermann, PhD

Throughout the drug discovery and development phases, analysis and characterization of the new chemical entity, its metabolites, and its degradation products require deployment of a number of analytical techniques. This chapter describes the principles, instrumentation, and application of some of the major techniques used in these analyses.

Discovery Research is responsible for identifying new chemical entities that may be useful for treating a disease or disorder. The various phases of drug discovery and development are shown in Figure 34-1. There are three phases to Hypothesis Generation: Target Identification and Validation, Assay Development, and Lead Generation.

The first process in Hypothesis Generation is Target Identification/Validation. A target may be a site in the human body that is linked in some manner to a disease, such as a specific gene or process within the body. A target may also be outside the human body, such as viruses, bacteria, or parasites. The task at this point is to first identify potential target-disease linkages. The next step is to test the target-disease linkage hypothesis in order to validate them using biological tools. Assay Development is the next step in the Hypothesis Generation process. The goal of Assay Development is to develop screen(s), which will allow for the identification of a lead compound. This involves designing the screen(s), and actually screening compounds against it. Given a target, one can design a test or screen(s) to see if a given chemical compound will have the desired biological effect. Next, we must be able to test a large volume of chemicals against the screen(s). These chemical compounds may be naturally occurring or they may be chemically synthesized. The last process in the Hypothesis Generation and testing stage is Lead Generation. The goal of Lead Generation is to find a molecule that causes a specific biological response: "a hit". At this point the primary concern is to identify those chemicals that had some level of desired effect in the screen. These chemicals then become leads. These leads serve as the starting point for further synthesis using chemistry tools in the next step in the overall process—Lead Optimization.

The lead compounds identified from the Hypothesis Generation stage must be refined and optimized prior to continuation of research, so that the value of research expenses are maximized. This process is called Lead Optimization, the first stage in the Development phase. Optimization is accomplished through refinement of in-vitro and in-vivo testing, performing necessary toxicology and pharmacokinetic studies, developing scale-up processes for drug development, and providing assay, stability and ADME (Absorption, Distribution, Metabolism, Excretion) testing and studies.

First Human Dose (FHD) Preparation includes a variety of protocols and studies that intend to prepare a particular compound for administration to healthy human volunteers. During this pre-clinical development, the new drug substance (NDS) is synthesized, analytically characterized and formulated for initial toxicology studies. Toxicology studies are designed to evaluate the effect of the compound on animals to determine what levels of the compound can be safely administered to humans. During this period, drug disposition scientists study the absorption, distribution, metabolism, and excretion in animals to further determine what the likely disposition of the compound will be in humans. Initial studies in humans are called Phase Ia clinical trials. In these studies, compounds are given to volunteers in very low single doses, then gradually escalated to determine tolerance to the substance. Although most studies are conducted in healthy, male volunteers, some drugs may be initially given to patients. If single doses of the experimental therapy are well tolerated, multiple doses will be administered to determine if the drug accumulates in body tissues and to further define the likely therapeutic level. An important objective of early clinical studies is to identify a pharmacodynamic effect relevant to the therapeutic application to guide the selection of doses for the treatment of patients.

Phase Ib and II are the phases where a drug's safety and efficacy are evaluated. With the First Efficacy Dose [FED], clinical investigators begin to study the effect of the drug candidate in a few patients to determine if a therapeutic response is produced. If a response is seen, the correct dose and regimen for the drug will be determined in Phase II clinical trials, involving several hundred patients. Concurrent with clinical trials, the chemistry and manufacturing components prepare larger quantities of material to support future clinical trials and to provide material for long-term toxicology studies. Commercialization is the final, and undeniably the most expensive portion of the Drug Development process. The Phase III clinical involves hundreds or thousands of patients, to fully evaluate efficacy and safety.

Finally, after approval is obtained, the product is launched, and it is here that the product's utility is truly demonstrated. Physicians are educated about the mechanism of action and dosage form and regimen. Physicians closely monitor the marketplace for additional adverse reactions that may appear in large-scale deployment. Throughout the life of the product, a primary responsibility of product development and manufacturing is to continually improve manufacturing processes and develop alternative drug delivery forms to retain a strong, competitive market position. This is the goal of Global Optimization.
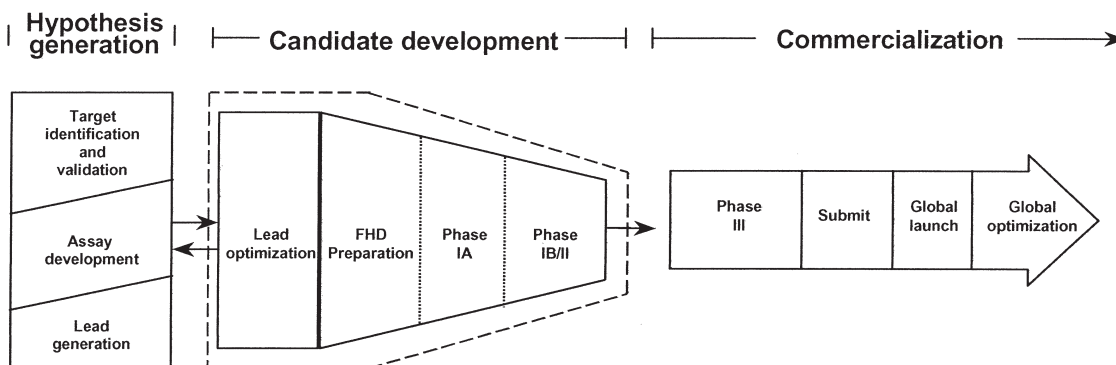
Figure 34-1. Drug development paradigm.

INSTRUMENTAL TECHNIQUES IN DRUG DISCOVERY AND DEVELOPMENT

Mass Spectrometry

Mass Spectrometry (MS) is an analytical spectroscopic tool primarily concerned with the separation of molecular (and atomic) species according to their mass. MS can be used in the analysis of many types of samples from elemental to large proteins and polymers. A mass spectrometer is an instrument that measures the masses of individual molecules that have been converted into ions, ie, molecules that have been electrically charged. The unit of mass is often referred to by chemists and biochemists as the Dalton (Da), and is defined as follows: 1 Da = (1/12) of the mass of a single atom of the isotope of carbon-12(^{12}C). This follows the accepted convention of defining the ^{12}C isotope as having exactly 12 mass units. A mass spectrometer measures the mass-to-charge ratio (m/z) of the ions formed from the molecules. The mass-to-charge ratio represents Daltons per fundamental unit of charge. In many cases, the ions encountered in mass spectrometry have just one charge (z = 1) so the m/z value is numerically equal to the molecular (ionic) mass in Da. Formation of gas phase samples ions is an essential prerequisite to the mass sorting and detection processes that occur in a mass spectrometer. Early mass spectrometers required a sample to be a gas, but due to recent developments, the applicability of mass spectrometry has been extended to include samples in liquid solutions or embedded in a solid matrix. The sample, which may be a solid, liquid, or vapor, enters the vacuum chamber through an inlet. Depending on the type of inlet and ionization techniques used, the sample may already exist as ions in solution, or it may be ionized in conjunction with its volatilization or by other methods in the ion source.

The gas phase ions are sorted in the mass analyzer according to their mass-to-charge (m/z) ratios and then collected by a detector. In the detector the ion flux is converted to a proportional electrical current. The data system records the magnitude of these electrical signals as a function of m/z and converts this information into a mass spectrum. A mass spectrum is a graph of ion intensity as a function of mass-to-charge ratio.

Some of the applications for mass spectrometry can be listed as follows:

- Accurate mass measurements can be used to match empirical formulae.
- Fragmentation fingerprints (specific to each compound) can be used to identify samples by comparison to fragment databases.
- Controlled fragmentation (through MS/MS and MS^n) can be used for structural elucidation of novel compounds.

- Common peaks observed in a spectrum can give useful information regarding functional groups.
- Relative isotope abundances are used to get information regarding the elements making up a compound.
- Complex mixtures can be analyzed via 'hyphenated' techniques such as GC-MS and HPLC-MS, thus negating the need for time-consuming sample purification.

MASS SPECTRA AND MOLECULAR STRUCTURE

With the low-energy electron beams in the order of 8 to 14 eV, it is possible to observe only the molecular ion (parent ion). Unlike other analytical methods, mass spectrometry gives an exact molecular weight. The mass spectrum of toluene shows a peak at $m/z = 92$ (m/z is the mass to charge ratio; for the parent ion this value also is the molecular weight), which is developed according to

$$\text{C}_6\text{H}_5\text{CH}_3 \xrightarrow{e} [\text{C}_6\text{H}_5\text{CH}_3]^+ + e \tag{17}$$
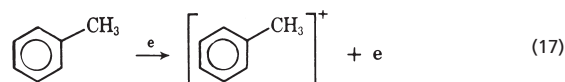
With a high-energy electron beam, in the order of 70 eV, the parent ion disintegrates, due to the removal of several electrons, giving positively charged and uncharged fragments. Adopting the symbolism for the transfer of a single electron by a single-headed arrow, two typical examples for fragmentation can be given by

$$\text{R}-\text{CH}_2-\text{NH}_2 \longrightarrow \text{R}^\bullet + \text{CH}_2=\overset{+}{\text{N}}\text{H}_2$$

$$\text{C}_6\text{H}_5-\overset{\text{O}}{\underset{\|}{\text{C}}}-\text{O}-\text{CH}_2-\text{CH}_2-\text{CH}=\text{CH}_2 \tag{18}$$

77 | 105 | 121 | 135 | 176
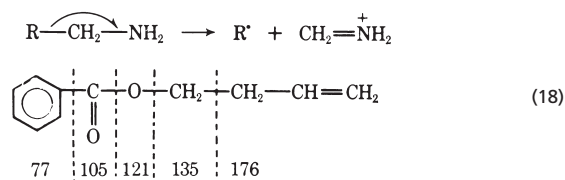
The mass spectrum of a compound, therefore, is a display of masses of molecular fragments together with the mass of the parent ion versus the relative abundance of each species as depicted by the peak heights.

The graphic form of the mass spectrum of toluene and its tabular presentation are depicted in Figure 34-2. The most intense mass peak is referred to as the *base peak* and is assigned an arbitrary value of 100; the other peaks are normalized relative to the base peak. Because the ratio of fragment abundance for a given compound remains constant, a mass spectrum (like an IR spectrum) becomes a *fingerprint* for each molecule.

m/e	% of base peak	m/e	% of P
38	4.4	92 (P)	100
39	16	93 (P + 1)	7.37
45	3.9	94 (P + 2)	0.29
50	6.3		
51	9.1		
62	4.1		
63	8.6		
65	11		
91	<u>100</u> (Base)		
92	68 (Parent)		
93	5.3 (P + 1)		
94	0.21 (P + 2)		
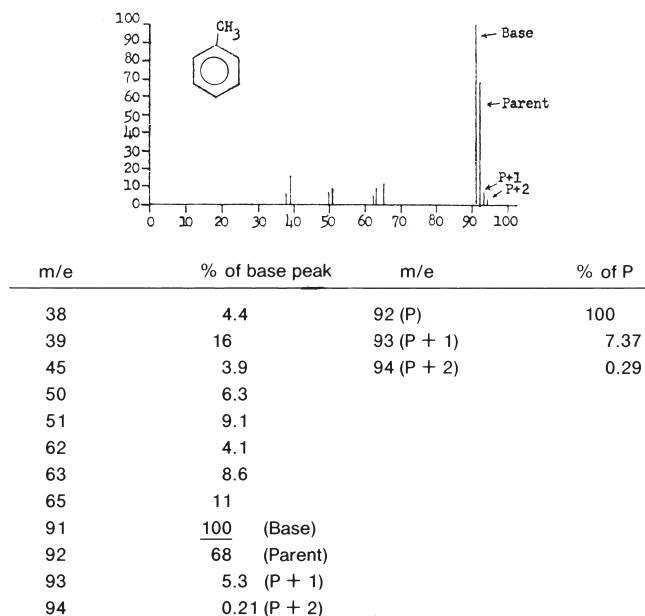
Figure 34-2. Relative abundance of various fragments shown in mass spectrum of toluene.

ISOTOPIC ABUNDANCE

The isotope abundance of atoms such as Cl, Br, S, and Si leads to the detection of these elements by mass spectrometry. For example, the ratio of ^{35}Cl to ^{37}Cl is 100 to 32.5.

For compounds of the general formula $C_wH_xN_yO_z$, contribution from the heavy isotopes can be calculated by

$$100\,\frac{P+1}{P} = 1.11w + 0.015x + 0.37y + 0.037z \qquad (1)$$

and

$$100\,\frac{P+2}{P} = 0.002wx + 0.004wy + 0.006w(w-1) + 0.20z \quad (2)$$

where P is the monoisotopic peak (parent peak; equivalent to the nominal molecular weight value) and $P + 1$ and $P + 2$ are the monoisotopic mass number plus one and two mass numbers, respectively. The relative isotope abundance of the heavy isotopes of each element determines the height of the $P + 1$ and $P + 2$ peaks. By consulting special tables of abundance factors for the $P + 1$ and $P + 2$ peaks, it is possible to determine an exact molecular formula from mass spectral data.

The following example represents the use of isotopic contribution in structural elucidation.

A compound with a mass spectrum of $P = 110$ (100%), $P + 1 = 111$ (5.5%), and $P + 2 = 112$ (0.3%) could be sorted out of the following molecular formulas with the molecular weight of 110:

Formula	P + 1	P + 2
$C_3H_2N_4O$	4.84	0.30
$C_4H_2N_2O_2$	5.20	0.51
$C_4H_4N_3O$	5.57	0.33
$C_4H_6N_4$	5.94	0.15
$C_5H_2O_3$	5.55	0.73
$C_5H_4NO_2$	5.93	0.55

The data reveal that the molecular formula of the compound is $C_4H_4N_3O$.

Table 34-1. Isotopic Abundances of Some Common Elements

ELEMENT	P MASS	%	P+1 MASS	%	P+2 MASS	%
H	1	100	2	0.015		
C	12	100	13	1.1		
N	14	100	15	0.37		
O	16	100	17	0.04	18	0.2
F	19	100				
Si	28	100	29	5.1	30	3.4
P	31	100				
S	32	100	33	0.79	34	4.4
Cl	35	100			37	32.0
Br	79	100			81	97.3

The isotope abundance of atoms such as Cl, Br, S, and Si leads to the detection of these elements by mass spectrometry. For example, the ratio of ^{35}Cl to ^{37}Cl is 100 to 32.0. The isotopic abundance of some common elements is shown in Table 34-1. The higher isotopic abundance of halogens such as Cl or Br provides characteristic fingerprint to the mass spectrum of molecules containing these atoms. The mass spectrum of Cl_2 and Br_2 is shown in Figure 34-3.

This characteristic is further illustrated in mass spectrum of protonated molecular ions (MH^+) of the drug labetolol as shown in the Figure 34-4.

STRUCTURE ELUCIDATION BY FRAGMENTATION PATTERNS—The fragmentation patterns of a few representative chemical classes are illustrated below. More detailed information can be obtained by consulting reference books on mass spectrometry (see Bibliography).

Several empirical rules of molecular fragmentation are:

- Cyclic compounds show an intense parent peak and a peak at the mass number of the ring.
- Saturated cyclic compounds lose side chains at the α-carbon. The peaks resulting from the loss of two atoms from the ring is more intense than the peaks from the loss of one atom.
- In cyclic compounds containing a double bond next to the side chain, cleavage occurs at the bond β to the ring.
- In olefins, cleavage occurs β to the double bond.
- In compounds with heteroatoms, cleavage occurs at the bond β to the heteroatom.
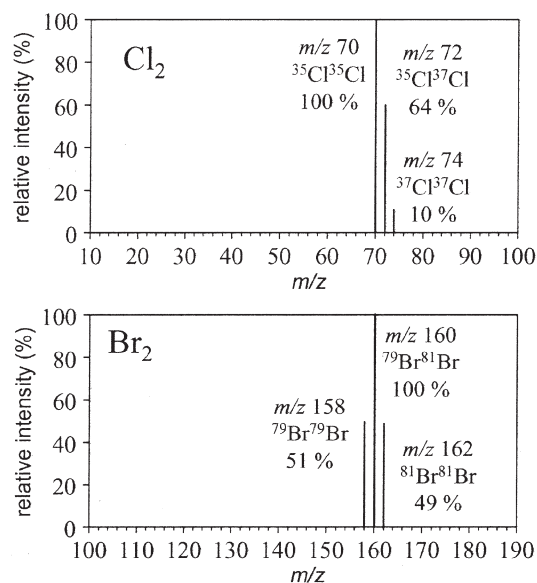
Figure 34-3. Mass spectrum of chlorine and bromine.
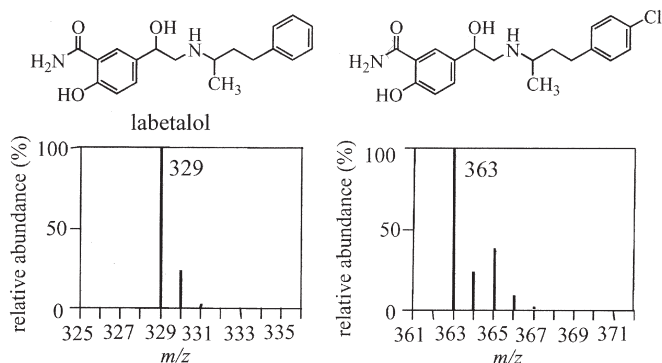
Figure 34-4. Mass spectrum: protonated molecular ions (MH+) of labetolol and its analog containing one chlorine atom.

- In hydrocarbon molecules, the ease of cleavage is in the following order: tertiary > secondary > primary. The positive charge remains on the branched fragment.
- In carbonyl containing compounds, cleavage occurs at this group, the positive charge remaining on the fragment containing the carbonyl function.

Molecular fragmentation may occur by one or a combination of the following processes: simple fission, simple rearrangement, complex fission, and complex rearrangement.

IONIZATION METHODS AND INSTRUMENTATION

The first step in producing the ions is the acquisition of energy by the molecules present in the sample. Some of the important methods of ionization are discussed below:

ELECTRON IMPACT (EI)—Electron impact ionization is the classical ionization technique in mass spectrometry. In the ion source ($10^{-7} - 10^{-5}$ mbar), the gaseous sample is bombarded with 70 eV electrons usually generated from a tungsten filament. Because the pressure is kept that low, ion-molecule reactions do not occur, eg, a $[M+H]^+$ signal due to proton transfer is not observed. The application of EI is restricted to thermally stable samples with low molecular masses (< ca. 2000 Da). Since the ion source temperature and the bombarding electron's energy are kept constant, the number and amount of fragments is constant for (almost) every mass spectrometer, too. Therefore, the number and amount of ionic fragments (*daughter ions*) and the amount of the M^+ is characteristic for each substance. Electron impact ionization has following characteristics:

- Can be used for GC/MS systems and direct inlet techniques.
- Produces "classical" compound spectra that are library searchable and/or interpretable.
- Useful for positive compound identification and/or structure elucidation.
- EI spectra are relatively easy to obtain.
- Comparatively rugged and sensitive ionization technique.
- Can be employed for analyzing air- and moisture-sensitive compounds.
- Analytes have to be vaporized - problems with thermal degradation.

CHEMICAL IONIZATION (CI)—Chemical Ionization is an ionization technique similar to the classical EI but the knowledge and results of ion-molecule reactions are exploited. In CI a similar ion source is used like in EI. One notable exception: The CI ion source is almost closed, ie, much smaller holes as the EI source, leading to high pressures (ca. 10^{-3} to 1 mbar). CI uses a reagent ion to react with the analyte molecules to form ions by either a proton or hydride transfer:

$$MH + C_2H_5^+ \rightarrow MH_2^+ + C_2H_4$$

$$MH + C_2H_5^+ \rightarrow M^+ + C_2H_6$$
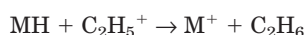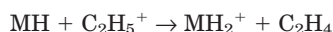
The reagent ions are produced by introducing a large excess of methane (relative to the analyte) into an electron impact (EI) ion source. Electron collisions produce CH_4^+ and CH_3^+ which further react with methane to form CH_5^+ and $C_2H_5^+$:

$$CH_4^+ + CH_4 \rightarrow CH_5^+ + CH_3$$

$$CH_3^+ + CH_4 \rightarrow C_2H_5^+ + H_2$$
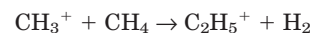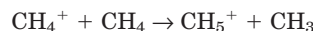
Sometimes other gases such as ammonia are also used as the reagent gas instead of methane.

Chemical ionization has following characteristics:

- Provides molecular weight information.
- Quantification is almost impossible without internal standards.
- CI can be used as ionization methods in GC/MS.

NEGATIVE-ION CHEMICAL IONIZATION (NCI)—Many important compounds of environmental or biological interest can produce negative ions under the right conditions. Negative ions can be produced by a number of processes. The electron energy is very low, and the specific energy required for electron capture depends on the molecular structure of the analyte. Benefits of NCI are efficient ionization, higher sensitivity and less fragmentation than positive-ion EI or CI. There is also a greater selectivity for certain environmentally or biologically important compounds. The limitations are that not all volatile compounds produce negative ions and a poor reproducibility of the measurements.

FAST-ATOM BOMBARDMENT (FAB)—In FAB a high-energy beam of netural atoms, typically Xe or Ar, strikes a solid sample causing desorption and ionization. It is used for large biological molecules that are difficult to get into the gas phase. The atomic beam is produced by accelerating ions from an ion source though a charge-exchange cell. The ions pick up an electron in collisions with neutral atoms to form a beam of high energy atoms. The FAB spectrum contains often only a few fragments and a signal for the pseudo molecular ion, eg, $[M+H]^+$, $[M+Na]^+$, adducts. This makes FAB useful for molecular weight determination. However, the low m/z region is crowded with signals resulting from the matrix. These matrix signals are not very reproducible. Therefore, spectra correction and interpretation is not easily accomplished.

MATRIX-ASSISTED LASER DESORPTION IONIZATION (MALDI)—MALDI is a method of vaporizing and ionizing large biological molecules such as proteins or DNA fragments. The biological molecules are dispersed in a solid matrix such as nicotinic or sinnapinic acid. A UV laser pulse ablates the matrix that carries some of the large molecules into the gas phase in an ionized form so they can be extracted into a mass spectrometer. MALDI allows determination of the molecular weight of molecules up to 500 kDa, routinely 5 to 100 kDa (polymers, biomolecules, complexes, enzymes), depending on the analyzer. The MALDI technique can be coupled with a time-of-flight analyzer (resolution and accuracy of the spectra are low but easy to handle and hence, most commonly used), quadrupole analyzer, ion traps or a fourier-transform mass spectrometer (expensive, difficult to handle, low dynamic range, but very accurate). Matrix-assisted laser desorption ionization has following characteristics:

- Soft ionization method provides molecular weight information.
- Suitable for analyzing very large bio- or synthetic polymers.
- Sensitivity depends strongly upon the analyte.
- Suitable for analyzing polar and even ionic compounds (e.g. metal complexes).
- Less fragmentation.
- Pulsed ionization technique, in contrast to EI, CI, FAB, ESI, and APCI.

ATMOSPHERIC PRESSURE IONIZATION (API)—API is used in conjunction with LC/MS techniques. The ions are formed at atmospheric pressure. It is a very soft ionization

technique leading in formation of predominantly molecular ion with little or no fragmentation. There are two common types of atmospheric pressure ionization:

Electrospray Ionization (ESI)—The ESI source consists of a very fine needle and a series of skimmers. A sample solution is sprayed into the source chamber to form droplets. The droplets carry charge when the exit the capillary and as the solvent vaporizes the droplets disappear leaving highly charged analyte molecules. Electrospray ionization is the method of choice for proteins, oligonucleotides and metal complexes. However, the sample must be soluble in low boiling solvents (acetonitrile, MeOH, CH_3Cl, water. . .) and stable at very low concentrations, ie, 10^{-2} mol/l. The characteristics of ESI can be summarized as follows:

- Soft ionization method provides molecular weight information.
- Suitable for analyzing large bio- or synthetic polymers.
- Sensitivity depends strongly upon the analyte.
- Suitable for analyzing polar and even ionic compounds (e.g. metal complexes).
- Less fragmentation.
- Enables LC / MS coupling.

Atmospheric Pressure Chemical Ionization (APCI)—Atmospheric pressure chemical ionization is closely related to ESI. The ion source is similar to the ESI ion source. In addition to the electrohydrodynamic spraying process, a corona-discharge needle at the end of the metal capillary creates a plasma. In this plasma proton transfer reactions and to a small amount fragmentation can occur.

Depending on the solvents, only quasi-molecular ions like $[M + H]^+$, $[M + Na]^+$ and M^+. (In the case of aromatics), and/or fragments can be produced. Multiply charged molecules $[M + nH]^{n+}$, as in ESI, are not observed. The characteristics of APCI can be summarized as follows:

- Provides molecular weight information.
- Sensitivity depends strongly upon the analyte.
- Suitable for analyzing less polar compounds compared to ESI.
- Increased fragmentation compared to ESI.
- Enables coupling MS and LC with flow rate up to 1 ml/min.

INSTRUMENTATION

The typical instrumentation for mass spectrometry consists of the following components:

SAMPLE INLET—The sample introduction systems produce vapors from the samples or reduce the pressure of the gaseous samples. Sample inlets include batch inlet for gases, and direct heated probes for solid samples. Hyphenated interphases with on-line sample introduction techniques such as gas or liquid chromatographic systems and electrophoresis systems are used extensively.

ION SOURCE—All ion sources produce analyte ions and introduce a suitable ion beam into the analyzer. Various ionization techniques to produce these ions are discussed above.

MASS ANALYZER—The center of any mass spectrometer is the mass selective analyzer. The main function of a mass analyzer is to resolve ions of the same m/z from all other ions and to focus the individual ion beams of discrete mass onto a detector or into a second ionization chamber or into a collision cell. The mass analyzers in current use can be classified in to two distinct classes. The first set of analyzers accomplish ion separation in a linear, *in space*, mode. These analyzers are: quadrupole, magnetic and time-of-flight. The second set of analyzers accomplishes ion separation *in time*. These analyzers are: ion trap and ion cyclotron resonance (Fourier transform).

In the *quadrupole mass analyzer,* four electric poles (a quadrupole) replace the magnetic field application procedure. The ions entering from the top, travel with a constant velocity in a direction parallel to the poles (Z direction) and acquire stable oscillation in the X and Y directions. This usually is accomplished by applying a dc voltage as well as radio frequency (rf) to the poles. Only one m/z ratio can pass through the quadrupole mass analyzer and be detected for a given rf potential and rf frequency. Therefore, a very rapid sweep can be performed by varying the rf frequency while rf and dc potentials are constant or *vice versa*. An advantage of the quadrupole is

that it does not require focusing slits and this results in higher sensitivity, as the resolution is only a function of the number of cycles an ion spends in the field.

In *time of flight spectrometers,* the ions of different mass are given the same kinetic energy allowing them to acquire different velocities, and have a time of flight that depends only on their mass—the lighter the ions, the faster they can travel through the field-free region. Hence, the original beam of ions tends to separate into several layers of ions, depending on their mass that bombarded the cathode of the ion-detector sequentially, and their transit time is calculated. Because a complete mass spectrum can be repeated 20,000 times in 1 second, the *time of flight* instrument is extremely useful in kinetic studies of fast reactions.

ION DETECTOR—The ions in each ion beam of different m/z are "counted" at a collector tube. This yields an analog signal that is amplified to provide electron currents that are representative of the number of collected ions of a particular m/z value.

COMPUTER SYSTEM—Digitized electrical data from the ion collector is fed into a computer where it can be processed further and simultaneously matched with a host of possibly identical spectra stored in the computer. This allows rapid accumulation, manipulation, and interpretation of the mass spectra. Dedicated computers are an integral part of a mass spectrometer. These perform a variety of functions such as automatic tuning, mass calibration, processing data etc. The advancements in information technology have helped in making mass spectrometers more powerful in terms of data acquisition as well as data processing.

VACUUM SYSTEM—All mass spectrometers operate under high vacuum that helps in eliminating unwanted collisions among the ions and between ions and neutral molecules. High vacuum also helps in preventing gas discharge from high voltages used in certain ion detectors. High vacuum also minimizes background and cross contamination between successive samples particularly in hyphenated systems.

HPLC coupled with API (ESI and APCI) MS techniques is one of the most widely used MS techniques to answer the bioanalytical issues during drug development. The usefulness and versatility of this technique will be illustrated in following section.

The production of ions by evaporation of charged droplets obtained through spraying or bubbling, has been known about for centuries, but it was only fairly recently discovered that these ions may hold more than one charge. A model for ion formation in ESI, containing the commonly accepted themes, is described below:

Large charged droplets are produced by *pneumatic nebulization*; ie, the forcing of the analyte solution through a needle (Fig 34-5) at the end of which is applied a potential. The potential used is sufficiently high to disperse the emerging solution into a very fine spray of charged droplets all at the same polarity. The solvent evaporates away, shrinking the droplet size and increasing the charge concentration at the droplet's surface. Eventually, at the Rayleigh limit, Coulombic repulsion overcomes the droplet's surface tension and the droplet explodes. This *Coulombic explosion* forms a series of smaller, lower charged droplets. The process of shrinking followed by explosion is repeated until individually charged *naked* analyte ions are formed. The charges are statistically distributed among the analyte's available charge sites, leading to the possible formation of multiply charged ions under the correct conditions. Increasing the rate of solvent evaporation, by introducing a drying gas flow counter current to the sprayed ions increases the extent of multiple-charging. Decreasing the capillary diameter and lowering the analyte solution flow rate, ie, in nanospray ionization, will create ions with higher m/z ratios (ie, it is a softer ionization technique) than those produced by *conventional* ESI and are of much more use in the field of bioanalysis. A positive ion LC/ESI/MS spectrum of the drug ganciclovir (MW 255) is shown in Figure 34-6. A strong signal
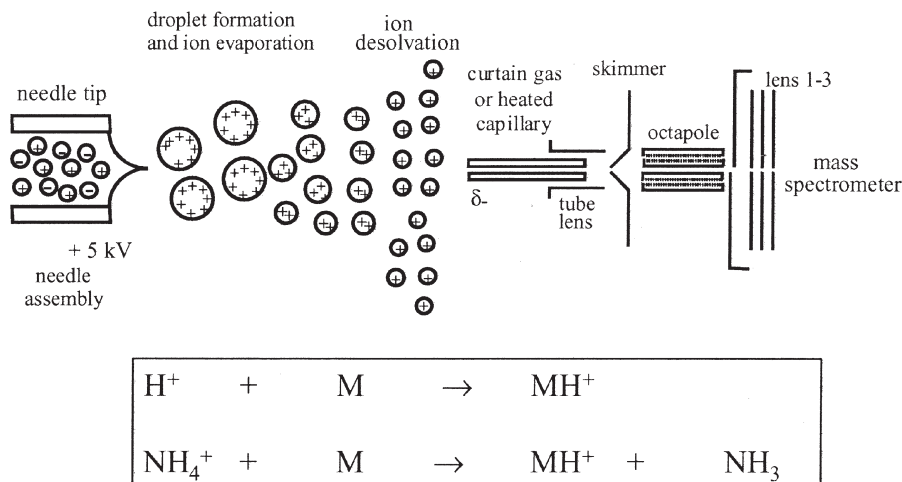
Figure 34-5. Positive electrospray ionization.

corresponding to protonated molecular $[MH]^+$ ion at m/z 256 is obtained. In negative ion LC/ESI/MS mode a strong signal corresponding to the loss of a proton $[MH]^-$ at m/z 254 is obtained.

TANDEM MASS SPECTROMETRY

Tandem mass spectrometry (MS/MS) is used to produce structural information about a compound by fragmenting specific sample ions inside the mass spectrometer and identifying the resulting fragment ions. This information can then be pieced together to generate structural information regarding the intact molecule. Tandem mass spectrometry also enables specific compounds to be detected in complex mixtures on account of their specific and characteristic fragmentation patterns.

A tandem mass spectrometer is a mass spectrometer that has more than one analyzer, in practice usually two. The two analysers are separated by a collision cell into which an inert gas (eg, argon, xenon) is admitted to collide with the selected sample ions and bring about their fragmentation. The analysers can be of the same or of different types, the most common combinations being:

quadrupole-quadrupole
magnetic sector–quadrupole
magnetic sector–magnetic sector
quadrupole–time-of-flight.

The basic modes of data acquisition for tandem mass spectrometry experiments are as follows:

PRODUCT OR DAUGHTER ION SCANNING—The first analyzer is used to select user-specified sample ions arising from a particular component; usually the molecular-related (ie, $(M+H)^+$ or $(M-H)^-$) ions. These chosen ions pass into the collision cell, are bombarded by the gas molecules which cause fragment ions to be formed, and these fragment ions are analyzed, ie, separated according to their mass to charge ratios, by the second analyser. All the fragment ions arise directly from the precursor ions specified in the experiment, and thus produce a fingerprint pattern specific to the compound under investigation. This type of experiment is particularly useful for providing structural information concerning small organic molecules. An example of product ion scanning is provided in Figure 34-7. ESI/MS analysis of the drug labetalol gives predominantly $[MH]^+$ ion at m/z 329. This ion upon collision-induced dissociation produces a series of fragment ions.

PRECURSOR OR PARENT ION SCANNING—The first analyser allows the transmission of all sample ions, while the second analyser is set to monitor specific fragment ions, which are generated by bombardment of the sample ions with the collision gas in the collision cell. This type of experiment is particularly useful for monitoring groups of compounds contained within a mixture, which fragment to produce common fragment ions, eg, aliphatic hydrocarbons in an oil sample, or glucuronide conjugates in urine.
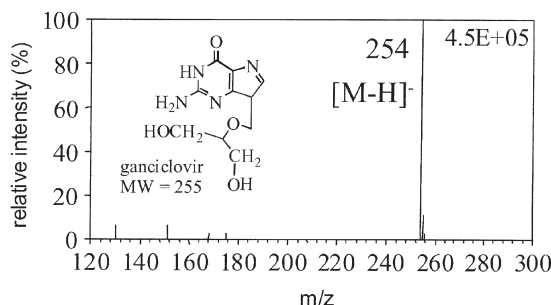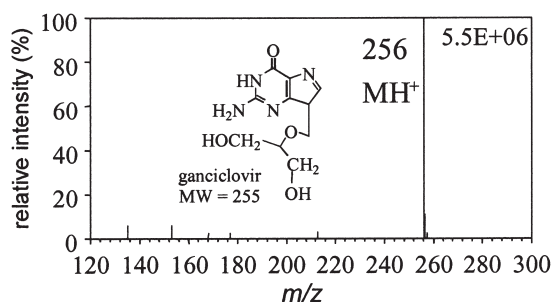
Figure 34-6. Positive (*upper panel*) and Negative (*lower panel*) LC/ESI/MS spectrum of ganciclovir.
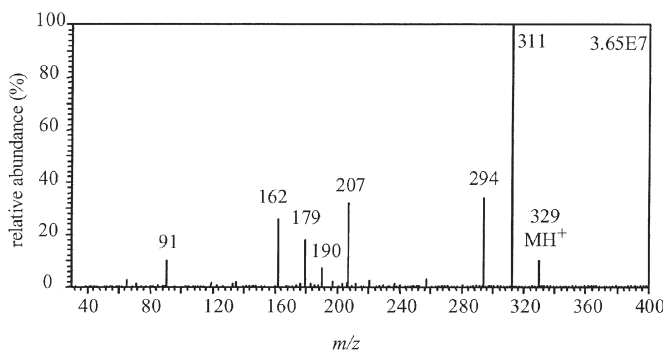
Figure 34-7. LC/ESI/MS/MS spectrum of labetalol.
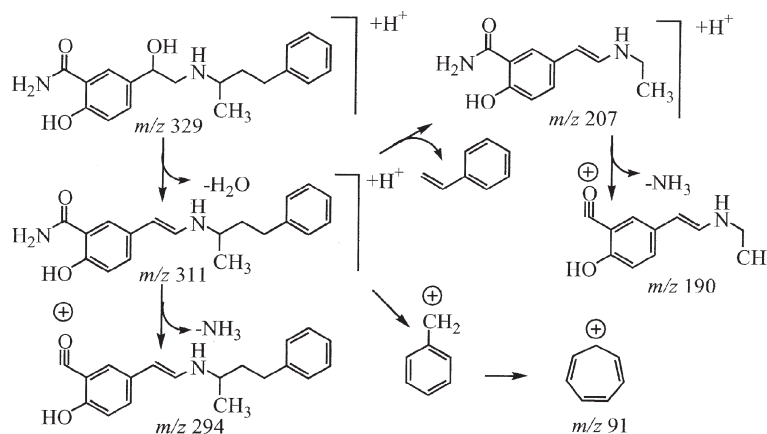
Figure 34-8. Product ions from labetelol.

CONSTANT NEUTRAL LOSS SCANNING—This involves both analysers scanning, or collecting data, across the whole m/z range, but the two are offset so that the second analyser allows only those ions, which differ by a certain number of mass units (equivalent to a neutral fragment) from the ions transmitted through the first analyser. This type of experiment could be used to monitor all of the carboxylic acids in a mixture. Carboxylic acids tend to fragment by losing a (neutral) molecule of carbon dioxide (CO_2), which is equivalent to a loss of 44 Da or atomic mass units. All ions pass through the first analyser into the collision cell. The ions detected from the collision cell are those from which 44 Da have been lost.

SELECTED/MULTIPLE REACTION MONITORING (SRM/MRM)—Both of the analysers are static in this case as user-selected specific ions are transmitted through the first analyzer and the second analyzer measures user-selected specific fragments arising from these ions. The compound under scrutiny must be known and have been well characterised previously before this type of experiment is undertaken. This methodology is used to confirm unambiguously the presence of a compound in a matrix, eg, drug testing with blood or urine samples. It is not only a highly specific method but also has very high sensitivity.

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance spectroscopy (NMR) is one of the most powerful tools used for the elucidating the structure of organic molecules. As with other forms of spectroscopy it involves the absorption of electromagnetic radiation, in this case in the radio frequency range (4–900 MHz). NMR is based on the fact that atomic nuclei have quantized spin states, which may be differentiated in the presence of a strong magnetic field. The application of energy in the form of a radio frequency, orthogonal to the applied magnetic field induces transitions between the allowed states resulting in the absorption of energy. The corresponding frequency of the absorbed energy depends, not only on the nuclei in question (eg, ¹H), but also depends on its electronic environment. Absorbed frequencies are therefore influenced by nature of chemical bonding, an attribute that ultimately leads to the use of NMR as a structural tool.

This section will deal primarily with proton (¹H) and carbon-13 (¹³C) NMR, as these nuclei are the primary nuclei studied in organic chemistry. Following a brief discussion of NMR theory, sufficient background will be given to provide the reader with the fundamental knowledge of how NMR is used for structural elucidation. This discussion will include an examination of the major 2-D techniques used. The section will conclude with a brief overview of the instrumentation associated with FT-NMR along with a brief overview of important advances including magic angle spinning (MAS) for the analysis of solids, the use of pulsed field gradients (PFG), and the evolution in probe technology for applications such as LC-NMR.

NMR THEORY

As early as 1924, Pauli suggested that certain atomic nuclei could have spin as well as a magnetic moment from rotation around their axes. He theorized that in the presence of an external magnetic field, nuclei could be split in to different energy levels. In other words, nuclei could align themselves either with or opposed to the magnetic field, much like the phenomenon observed with a bar magnet. Experimental verification of these theoretical concepts, however, did not occur until 1946 when Bloch at Stanford and Purcell at Harvard independently demonstrated that nuclei absorb electromagnetic radiation in the presence of a strong magnetic field. The two physicists shared the Nobel Prize in physics in 1952 for their work.

Quantum mechanics can be used to describe the properties of spinning nuclei. Specifically, the angular momentum of the spinning charge is expressed by the *spin quantum number*, I. The spin quantum number can adopt integer or half-integer values expressed in units of h/2π, where h is Plank's constant. As shown in Table 34-2, spin numbers vary for various nuclei depending on the relationship between the number of protons and neutrons in the nucleus.

Any nuclei of I > 0 placed in a magnetic field will assume a maximum number of orientations equal to 2I +1. Because I = ½ for the proton, two orientations or spin states exist: aligned with field (low energy) and opposed to field (high energy). Pharmaceutical applications of NMR primarily focus on nuclei of I = ½ (Table 2). As stated previously, two of these nuclei (¹H and ¹³C) will be considered in the chapter.

The separation of the energy levels (spin states) is a function of the nuclear magnetic moment (μ), the external magnetic field (H_0) and the spin quantum number (I) according to equation 3.

$$E = \mu H_0 / I \qquad (3)$$

In the presence of a strong magnetic field, nuclei will precess about an axis parallel to the magnetic field. This process is depicted for a single nucleus in Figure 34-9.

Table 34-2. Spin Quantum Numbers for Various Nuclei

SUM OF PROTONS PLUS NEUTRONS	SPIN QUANTUM NUMBER (I)	EXAMPLES OF NUCLEI
Even	0	¹²C, ¹⁶O, ³²S
Odd	1/2	¹H, ¹³C, ¹⁹F, ³¹P
Odd	3/2	¹²⁷I, ¹¹B, ⁷⁹Br
Even	1	²H, ¹⁴N
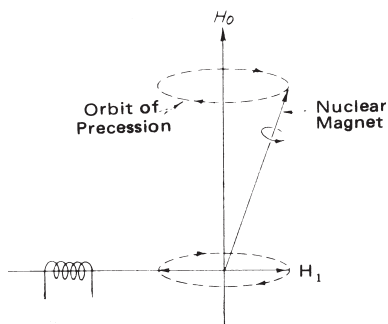
Figure 34-9. The spinning and precessing of a nuclear magnet in an external magnetic field.

The angular velocity of a precessing nucleus, ω_0, is related to the applied magnetic field according to:

$$\omega_0 = \gamma H_0 \tag{4}$$

γ is the magnetogyric ratio (a constant for a given nucleus which is directly proportional to the magnetic moment). The greater the angular velocity, the greater the energy required to flip a nucleus from its low energy state (aligned with H_0) to its high energy state (opposed to H_0). During a NMR experiment, a sample is placed in a homogeneous magnetic field (H_0) and an external rf oscillator coil is used to introduce a second magnetic field orthogonal to H_0. When the applied rf frequency (ν) is equal to $\omega_0/2\pi$, a nuclei is able to absorb energy and flip to the excited state. As this occurs, a nucleus is said to be in "resonance" and the associated frequency is referred to as the resonance frequency. The energy absorbed in this transition induces a potential in a receiving coil placed orthogonal to both H_0 and the magnetic field of the oscillating coil. In older generation NMR instruments, referred to as continuous wave (CW), the applied frequency is held constant while H_0 is scanned. As different nuclei are consecutively brought into resonance, a voltage is induced proportional to the number of nuclei present in the sample. The NMR spectrum produced by this process is essentially a plot of applied frequency (ν) versus absorbed peak intensity. In this spectrum, peak intensity is expressed in arbitrary units; however, this signal reflects the concentration of the sample as well as the relative abundance of each type of magnetically distinct nucleus that exists in the molecule analyzed.

Fortunately, the resonance frequency of a particular nucleus (eg, ^1H) varies considerably according to its electromagnetic environment. Since this environment is strongly influenced by chemical structure, NMR is a powerful tool for the structural assignment of organic molecules. The specific correlation between resonance frequency and chemical structure is referred to by the term *chemical shift*. This concept will be discussed subsequently in more detail when we address the use of NMR for structure determination.

ABSORPTION AND RELAXATION

An interesting distinction about NMR compared to other forms of spectroscopy is that there are roughly equal populations among the higher and lower energy states. Actually, a slight excess of nuclei exists in the lower energy state. In the absence of an applied magnetic field the population of the two states is governed by a Boltzmann distribution. In most other forms of spectroscopy, the lowest energy state (ground state) is heavily populated relative to the excited states. For absorption of energy to occur, there must be an excess population in the lower state. However, because this excess in NMR is

slight, the phenomenon of signal saturation occurs readily as the lower state is depleted (ie, the two states achieve equal populations). This phenomenon also accounts for the relative insensitivity of NMR compared to other forms of spectroscopy.

The population in the upper state also depends on a phenomenon known as relaxation, which refers to any process that removes nuclei from an excited state. Relaxation is also important to NMR peak width, since the width of an absorption band is inversely proportional to the lifetime of the excited state. Two basic relaxation mechanisms are operative in NMR: spin-spin relaxation and spin-lattice relaxation.

SPIN-SPIN (TRANSVERSE) RELAXATION—Spin-spin relaxation involves the mutual exchange of energy between two proximal precessing nuclei. Basically, when two neighboring nuclei have identical precession rates, but exist in different spin states, the magnetic fields of each nucleus can interact to cause a mutual exchange in spin-states (low to high and high to low). Obviously, this type of relaxation does not help maintain an excess lower spin state population. It can, however, increase line broadening by decreasing the average lifetime of a nucleus in excited state.

SPIN-LATTICE (LONGITUDINAL) RELAXATION—Spin-Lattice relaxation involves the transfer of energy to lattice components (surrounding molecules) as nuclei return from the higher to a lower spin state. Energy dissipated to the lattice increases the vibrational, translational and rotational energy of adjacent nuclei. Because this process replenishes the lower state, it helps alleviate signal saturation. The extremely broad NMR signals observed for solids or viscous liquids are attributed in part to a limited facility for spin-lattice relaxation.

INTERPRETATION OF SPECTRA—To begin to understand how NMR is used for structural interpretation, it is important to introduce specific terms or concepts such as shielding, chemical shift, anisotropy and spin-spin coupling. Each of these terms will be described below in the context of proton NMR.

CHEMICAL SHIFT—The magnetic environment of any nucleus is influenced by circulating electrons, since electrons are capable of producing small magnetic fields. When these fields counteract the applied magnetic field (H_0), the nucleus is said to be shielded. A stronger magnetic field must therefore be applied to transition protons shielded by circulating electrons. As one might expect, shielding can be reduced by the presence of electron-withdrawing substituents, such as oxygen or halogens. Consequently, protons attached or adjacent to electronegative substituents require less magnetic field strength to achieve resonance and are said to be displaced downfield or deshielded.

Chemical shift is the term used to describe the unique magnetic field strength required to achieve resonance for any given proton. As discussed earlier, the required magnetic field for resonance (and thus chemical shift) is influenced by the electromagnetic environment of the proton. To correlate observed chemical shifts to chemical structure, it is important to define chemical shift in a manner that is independent of instrument or applied field strength. By convention, tetramethylsilane (TMS) is used as a reference in proton NMR and is assigned a chemical shift of zero. TMS was selected because its methyl protons are found well upfield from almost all known proton resonances owing to the electropositive nature of silicon (maximum shielding). Chemical shifts are defined by the following equation and are expressed in parts per million using the symbol δ.

$$\delta = (\nu_s - \nu_{TMS}/\nu_{rf}) \times 10^6 \tag{5}$$

In equation (5), ν_s and ν_{TMS} are the field strengths (in Hz) for the sample and reference, respectively, while ν_{rf} refers to the frequency of the applied rf signal. Note, the reader should also be made aware of a second convention using the symbol τ (where $\tau = 10 - \delta$).

As δ (ppm) increases, the proton is said to be shifted downfield (a higher frequency/lower field required for resonance). For the purpose of illustration, the chemical shifts (ppm) for the methyl protons of the molecular series CH_3X are as follows: CH_3I (2.16), CH_3Br (2.68), CH_3Cl (3.05) and CH_3F (4.26). In this series the effect of electronegativity is clear as the most electronegative element fluorine creates the highest deshielding.

Chemical shifts are influenced by additional sources of magnetic fields correlated to chemical structure. Several cases involve the participation of π electrons. A classic illustration is the apparent anomaly in the chemical shift (δ) of acetylenic (2.35) and olefenic (4.60) protons. Since most methylenic protons have chemical shifts below 2 ppm, one might expect that the deshielding caused by the introduction of π electrons would result in higher deshielding for acetylene than ethylene. This phenomenon is explained by the illustration in Figure 34-10a and 34-10b where the magnetic field induced by the triple bond of acetylene acts in concert with H_0. This effect, referred to as *diamagentic anisotropy*, also accounts for the deshielding of the aromatic protons of the phenyl group (S 7–8 ppm) as well as the extreme deshielding observed for aldehyde protons (Fig 10b)

SPIN-SPIN COUPLING—Figure 34-11a depicts the NMR spectrum of ethanol. In this spectrum one might expect to observe three distinct peaks for CH_3, CH_2 and OH in the abundance of 3:2:1. Interestingly, this is not the case. Although three clusters of peaks are observed, two occur as multiplets. These multiplets are formed by a phenomenon called spin-spin coupling (often referred to as *splitting* or scalar coupling), which is caused by the influence of the spin states of neighboring protons transmitted through chemical bonds.

Figure 34-11b illustrates the probable nuclear arrangement of the —CH_2— and —CH_3 groups of ethanol. The spin coupling which leads to the observed multiplicity is sometimes called *3-bond* coupling since information about neighboring spin states is transmitted through three chemical bonds (ie, 2 C-H and 1 C-C). This means that the multiplicity observed in 1H-NMR spectra is due to the influence of protons on adjacent carbon atoms. The formula for spin-spin generated multiplicity is $2nI + 1$, where n is the number of equivalent nuclei of spin I. For ^1H and ^{13}C, this formula may be re-written as $n + 1$. In Figure 34-11a, the multiplet for the —CH_2— group is a quartet of peaks of the intensity ratio 1:3:3:1. The —CH_3 multiplet occurs as a triplet of intensity ratio 1:2:1. An explanation of this splitting pattern is shown in Figure 34-11b, which gives the possible combinations for parallel and antiparallel spin for —CH_3 (upper) and —CH_2- (lower). Note, this illustration is consistent with the $n + 1$ rule. In addition, the complete integration for the two multiplets gives a ratio of 3:2 as expected from the overall proton abundance of the two groups.
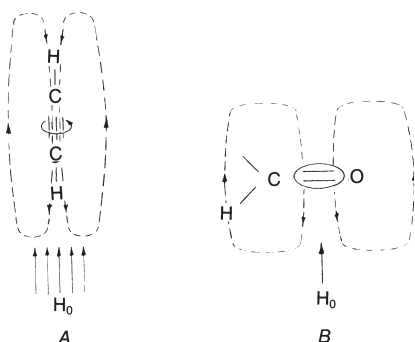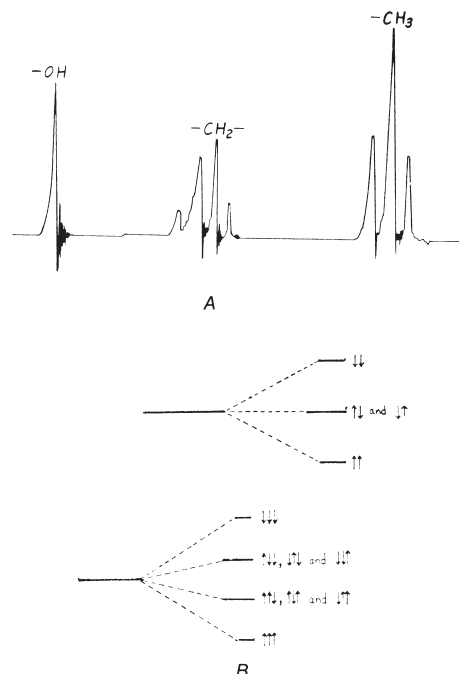
Figure 34-11. **A.** A high-resolution NMR spectrum of ethanol. **B.** Spin-spin splitting presentation of proximal CH_3 and CH_2.

Although spin-spin coupling adds complexity and reduces the sensitivity of NMR, it is an essential feature for organic structure elucidation because it allows the features in a NMR spectrum to be correlated to adjacent positions within a molecule. The distance between adjacent peaks within a multiplet is also diagnostic and is referred to as the scalar (through bond) coupling constant J. J values are on the order of a few Hz (rarely exceed 20 Hz). Coupling constants are generally denoted as J_{A-B}, where A and B refer to the corresponding positions on the molecule.

^1H-NMR SPECTRUM OF PROPRANOLOL—Several features of spectral interpretation are illustrated by the proton NMR spectrum of the common β-blocker, propranolol. Propranolol has 16 carbons as well as 21 hydrogens as labeled in Figure 34-12. Propranolol also has three heteroatoms, two of which have exchangeable hydrogens. Because this NMR spectrum was acquired in a deuterated solvent (CD_3OD), these active hydrogens were readily exchanged by deuterium and are not observed in the NMR spectrum shown. Despite the use of deuterated solvents, the presence of the residual water in the sample is evident from the large water resonance at 4.83 ppm.

An illustration of spin-spin coupling is found in the three-carbon spin system defined by the isopropyl group. Figure 34-13a shows an expanded view of the multiplet centered at 3.49 ppm. This multiplet, which contains seven peaks, has a combined integration of 1.00 meaning that it represents a single proton. According to the $n+1$ rule, there must be a total of six protons on the neighboring carbon atoms, a situation that can only be explained by the methyl groups of the isopropyl moiety. Therefore, this peak is assigned to the methine proton of the isopropyl group (carbon 14).

The corresponding multiplet for the methyl groups (Fig 34-13b) is centered at 1.38 ppm and has a combined peak integration of 6. The expanded view of this multiplet reveals what appears to be a triplet, but is in reality an unresolved pair of doublets, each corresponding to one of the methyl groups. Because of the chiral center, these methyl groups are chemically nonequivalent and thus have slightly different chemical shifts. Each peak is split into a doublet by the methine proton.

Figure 34-10. The electron-induced magnetic lines of force. **A.** Shielded acetylenic proton. **B.** Deshielded aldehyde proton.
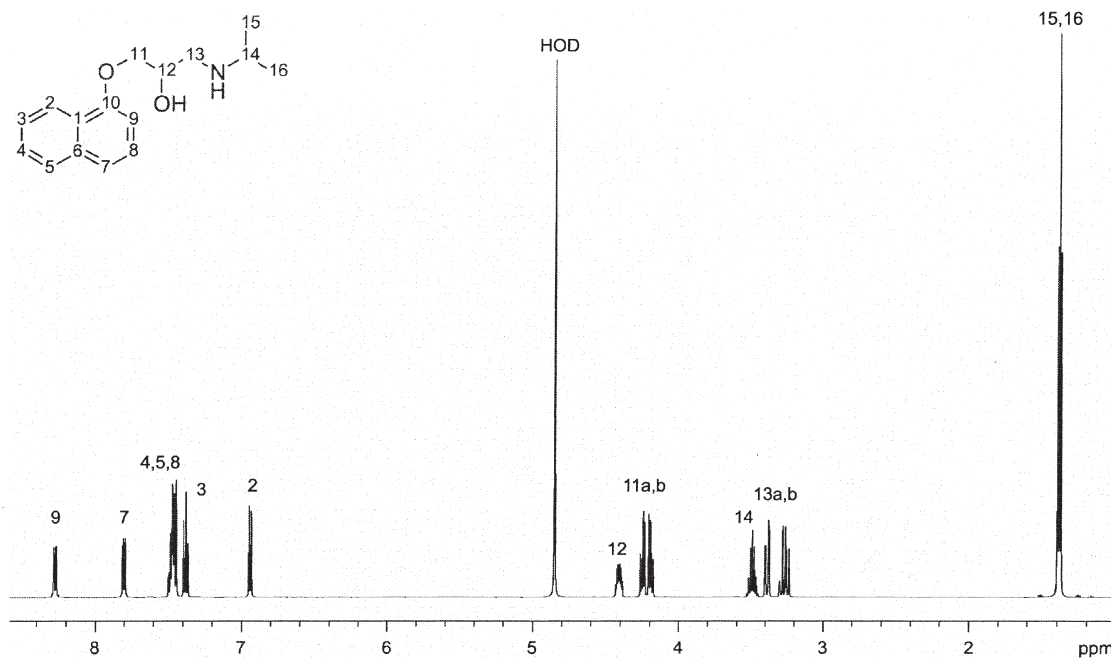
Figure 34-12. 1H-NMR spectrum of propranolol.

Table 34-3 provides a list of assignments for all 19 non-exchangeable protons in the ^{1}H-NMR spectrum of propranolol.

The assignments for propranolol were based on three pieces of information: chemical shift, peak integration and scalar coupling. Despite the complexity of the spectrum, the molecule can be broken down into 3 spin systems. The first system, corresponding to the isopropyl group, was already described. A second spin system is created by the interaction of the protons on carbons 11, 12, and 13. Multiplets for the protons on these carbons are centered at 4.22, 4.41 and 3.33, respectively. The proton on carbon 12 is readily identified since it is the most deshielded (attached to a primary alcohol) and has a peak integration of 1.00. In addition, this resonance exists as complex multiplet created from splitting by two adjacent methylene groups. One would predict that the protons on the methylene groups adjacent to carbon 12 would exist as doublets. This is true to a first approximation; however, additional splitting occurred resulting in the production of a doublet of doublets in each case (Fig 34-12). The proton pairs on carbons 11 and 13 are differentiated with the letters "A" and "B" in Table 34-3 indicating that they are not magneti-

cally equivalent. These sets of protons are said to be *diastereotropic*, a phenomenon introduced by the chiral site in propranolol. The assignments for the methylene protons on carbons 11 and 13 are consistent with predicted chemical shifts and peak integration across the complex doublets observed in each case.

The third system corresponds to the aromatic protons on the napthyl ring. Due to anisotropic effects the aromatic protons appear in the region from 6.5 to 8.5 ppm. Based on the information present in this region of the spectrum, it is possible to deduce the position of substitution on the napthyl ring. The reader is referred to Table 34-3 for a list of the assigned resonances in this region.

A technique frequently used in concert with ^{1}H-NMR is selective decoupling to identify the protons participating in a given spin system. Decoupling can be accomplished by irradiating the sample at the resonance frequency of one of the known protons. When this occurs, any peaks coupled to the proton in question automatically become decoupled causing any related multiplets collapse into a single peak. Historically, this process was repeated manually for each resonance, but for years has oc-
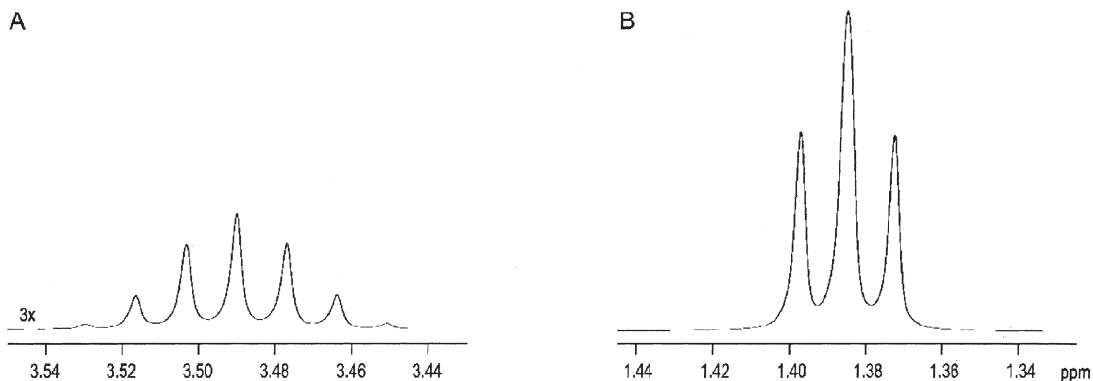
Figure 34-13. Expanded view of the muliplets related to the isopropyl moiety of propranolol. Figure 34-13a corresponds to the methine proton (carbon 14) and Figure 13b represents the methyl protons (carbons 14 and 15).

Table 34-3. Peak Assignments for the ^{1}H-NMR Spectrum of Propranolol

CHEMICAL SHIFT (PPM)	MULTIPLICITY	INTEGRATION (NO. OF PROTONS)	ASSIGNMENT (ATTACHED CARBON)
1.38	triplet (unresolved pair of doublets)	6	15, 16
3.26	doublet of doublets	1	13A
3.39	doublet of doublets	1	13B
3.49	septet	1	14
4.19	doublet of doublets	1	11A
4.25	doublet of doublets	1	11B
4.41	multiplet	1	12
4.83	singlet	n/a	residual water
6.94	doublet	1	2
7.38	triplet	1	3
7.47	multiplet	3	4, 5, 8
7.81	doublet	1	7
8.28	doublet	1	9

curred as an automated process using 2-D NMR sequences. An illustration of this technique is provided in the subsequent discussion of 2-D methods.

NMR Instrumentation
HISTORICAL OVERVIEW

The use of NMR as a tool for organic structure elucidation is now almost fifty years old. Over this period a number of important advances in instrumentation have occurred. The use of NMR as a structural tool began in the 1950s when the notion of chemical shifts and the recognition of spin-spin coupling occurred. Advances during the 1960s included the application of Fourier transformation (FT), the use of signal averaging for improved sensitivity and the first utilization of the Nuclear Overhauser Effect (NOE) for structural determination. In the 1970s, NMR became a more widely adopted technique. It was during this decade that FT-NMR became a commercial reality along with the introduction of superconducting magnets and computer controlled instrumentation. The introduction of magic angle spinning for solids was also introduced. Many of the advances in the 1980s were related to the implementation of 2-D techniques as well as larger sized magnets to enable more complex structural elucidation, such as the analysis of peptides and proteins. These trends continued in the 1990s along with the routine implementation of pulsed field gradients, which have improved the overall data quality obtained by 2-D techniques. The 1990s also witnessed several advances in probe technology, including the introduction of a viable interface between HPLC and NMR. A high level overview of the important trends that have led to the capabilities associated with modern NMR appears in the remainder of this section. Greater detail may be obtained from the sources cited in the bibliography.

FT-NMR

One of the chief limitations of NMR is sensitivity. This is particularly true for ^{13}C spectra which are about 6000-fold less intense that ^{1}H owing to a lower natural isotopic abundance and a weak magnetogyric ratio (γ). Although sensitivity issues still exist, NMR was revolutionized by the introduction of Fourier transform (FT) techniques. This technology has led to vastly improved sensitivity and has led to the introduction of 2-D NMR methods.

Prior to FT-NMR, data were acquired in a linear or scanned function. That is, the magnet was scanned over the range of desired chemical shift allowing nuclei to come into resonance in a consecutive fashion. In FT-NMR, data acquisition occurs in a multiplexed format, initiated by pulsing the sample with a 1–10

μsec pulse of radio frequency radiation encompassing multiple wavelengths (white noise). This step causes all protons to achieve resonance simultaneously. The induced voltage pattern detected from this process represents a complex time-domain signal known as the free induction decay (FID).

As depicted in Figure 34-14 the FID represents a complex pattern having a tapered shape. The fall-off in the FID signal is the result of spin-relaxation, whereas the complexity can be understood from the wealth of information encoded in a single FID transient. Obviously, the FID is of little use in its time-domain form. The act of Fourier transformation converts this time-domain signal into a frequency-domain pattern containing all information needed to reconstruct the complete NMR spectrum.

Typically, several FID transients are summed to produce a NMR spectrum. Signal averaging is commonly used to increase the signal-to-noise ratio in accordance with the Felgett advantage. With current technology, a complete ^{1}H NMR spectrum may be acquired using as little as 100 μg of material; however, low milligram quantities are typically used. For ^{13}C NMR spectra, several milligrams are recommended and data are often acquired over several hours (eg, overnight acquisition).

2-D NMR SPECTROSCOPY

To mange the complexity of NMR spectra, two-dimensional methods were introduced.[1] The term 2-D NMR refers to the use of a variety of pulse sequences that allow a sample to be perturbed along two independent time domains. Using 2-D methods, NMR spectra are no longer linear; rather they express
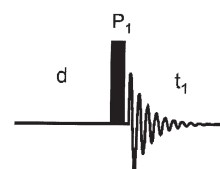
Figure 34-14. Pulse sequence for 1D 1H NMR data acquisition. The sample is irradiated with an rf-pulse (indicated as P1) containing multiple frequencies at an angle 90º with respect to the applied magnetic field. Following absorption of radiation, data acquisition occurs during the period shown as t1, to produce a time domain signal known as the free induction decay (FID). Fourier transformation is subsequently applied to derive frequencies corresponding to chemical shifts. The equilibration time between scans (inter scan delay) is depicted at the front of the sequence by the letter "d."

2D ^{1}H-^{1}H COSY
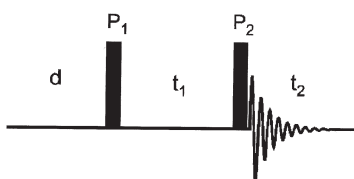
Figure 34-15. Pulse sequence for 2D 1H-1H NMR data acquisition by COSY. This sequence builds on the sequence 1D sequence shown in Figure 34-14 by inserting a second orthogonal rf-pulse (P2) prior to data acquisition, indicated as t2. To acquire a COSY spectrum, the evolution time between pulses P1 and P2 (t_1) is varied so that consecutive FIDs are acquired under different evolution times. This process ultimately yields a 2D contour map allowing scalar coupling interactions to be identified from cross-peaks in the 2D spectrum.

a matrix of all combinations of the two variables studied. Fourier transformation of each encoded time domain in the resulting FID yields a spectrum having two separate frequency axes. The 2-D spectrum is expressed in a plane defined by the two frequency axes with peak intensity displayed using a contour format.

The most widely used 2-D technique is named COSY, which stands for Correlation Spectroscopy. A generic pulse sequence for COSY appears in Figure 34-15. Comparison to the sequence in Figure 34-15 reveals the application of a second pulse along with the insertion of an additional time element shown in this Figure at t_1. In practical terms, COSY introduces a second decoupling dimension to a conventional ^{1}H-NMR allowing facile determination of all spin-systems present in a molecule.

Figure 34-16 displays the double quantum-filtered (DQF-COSY) spectrum of propranolol. The chemical shift region from 1 to 5 ppm is plotted orthogonally yielding a 2-D contour plot where the conventional 1-D ^{1}H-NMR spectrum is encoded along the diagonal of this plot. Peaks that fall off of the diagonal are referred to as "cross-peaks" because they allow two related portions of the NMR spectrum to be correlated. An ex-
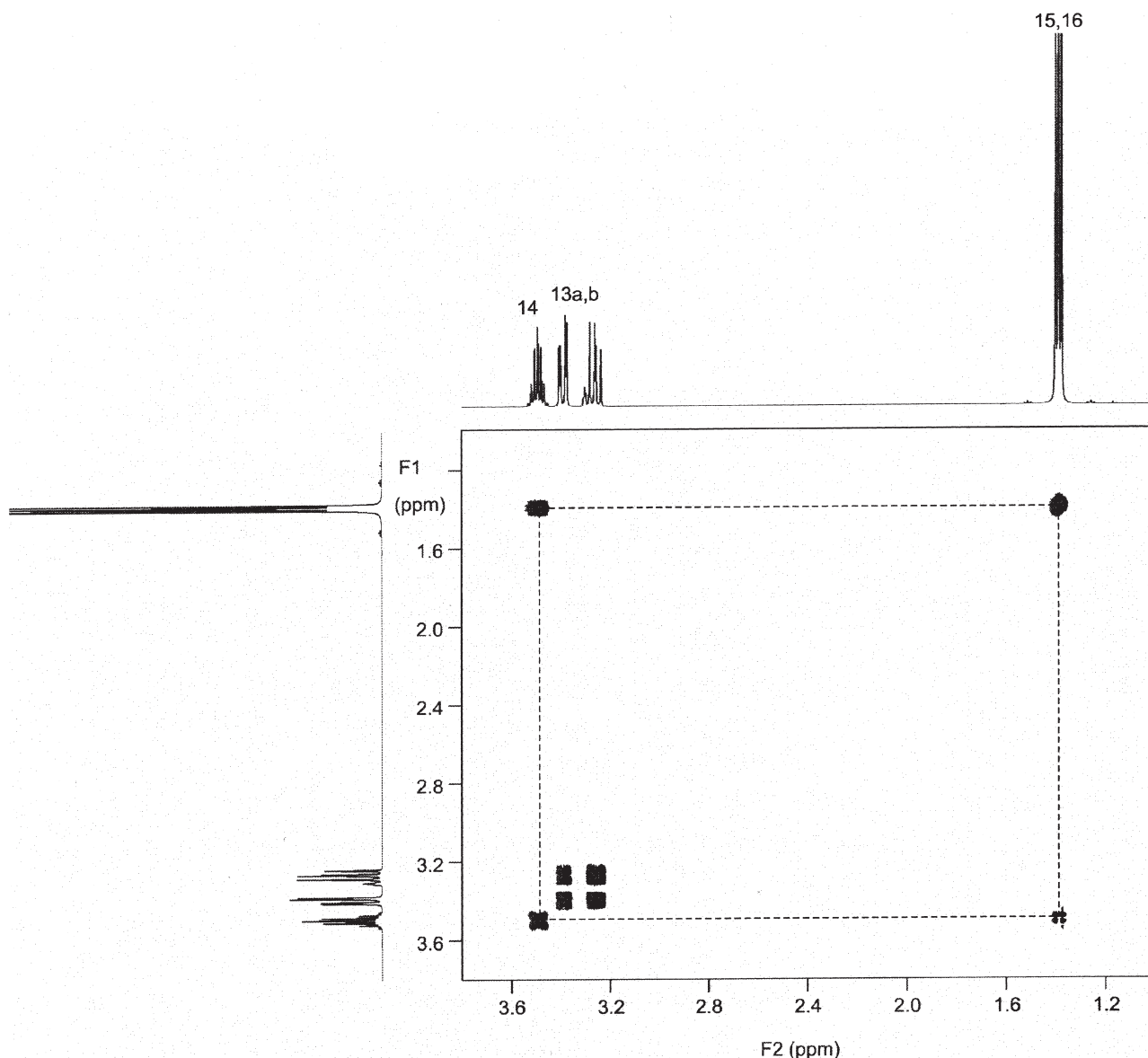
Figure 34-16. Region of the double quantum filtered COSY 2-D NMR spectrum of propranolol. The dotted lines identify protons that are spin-coupled, as indicated by the presence of cross-peaks which fall off the central diagonal. The resonances indicated correspond to the isopropyl spin system.

Table 34-4. Commonly Used 2-D NMR Methods

EXPERIMENT	FULL NAME	NUCLEAR CORRELATION	INFORMATION
COSY	Correlation Spectroscopy	^{1}H-^{1}H Homonuclear	Assignment of spin systems and J_{A-B} (scalar coupling constants)
NOESY	Nuclear Overhauser Enhancement Spectroscopy	^{1}H-^{1}H Homonuclear	Short range through space interactions (dipolar coupling)
HSQC	Heteronuclear Single Quantum Coherence	^{13}C-^{1}H Heteronuclear	An indirect method which allows gives carbon chemical shifts and single bondC-H connectivity
HMBC	Heteronuclear Multiple Bond Correlation	^{13}C-^{1}H Heteronuclear	Similar to HSQC, but gives information about C-H bond connectivity through as many as 4 bonds (long range).

ample illustrated in the Figure is the isopropyl spin-system discussed earlier. In this example the methine proton of carbon 14 (3.49 ppm) is shown to be spin-coupled to the methyl protons of carbons 15 and 16 by the perpendicular lines drawn on the spectrum connecting the resonances to the cross peak of interest. Application of the double quantum filtered variation of COSY has become quite popular since the cross peaks are split into finely divided multiplets providing information on the peak degeneracy produced by the associated spin-system. In addition, it is possible to derive coupling constants from the distance between adjacent peaks.

COSY is but one a several 2-D methods routinely used by NMR spectroscopists. Table 34-4 lists the most common 2-D methods and provides general information about their use. The techniques listed are classified as "homonuclear" meaning that common nuclei are involved or "heteronuclear" if the nuclei are different (eg, ^{1}H and ^{13}C). Heteronuclear methods are often "indirect" meaning the information is obtained indirectly about carbon atoms from their coupling to protons (improved sensitivity since ^{13}C not measured directly). The NOESY technique (based on the Nuclear Overhauser Effect) is particularly useful for distinguishing isomeric configurations since the signals involved arise from through space interactions (as opposed to through bond interactions).

PULSED FIELD GRADIENTS

In 1973 Lauterbur discovered that the use of linear gradients in applied magnetic field enable the determination of spatial position within the magnet.[2] This discovery ultimately led to the present day application of NMR for magnetic resonance imaging (MRI). In early 1990s, it was discovered that pulsed field gradients (PFG) could be used in conjunction with 2-D methods to improve overall data quality.[3]

Without going into detail, there are three basic benefits that have been derived through the application of PFGs. The first is that PFGs allow NMR spectra to be acquired without phase cycling. Phase cycling involves the acquisition of different types of data in alternate scans allowing the desired data to be combined from each scan. PFGs remove the considerable overhead associated with phase cycling resulting in faster data acquisition. The application of PFGs also leads to cleaner spectra by eliminating artifact peaks associated with phase cycling. Finally, PFGs help suppress the unwanted effects of solvent (eg, H_2O) in NMR spectra. Because of these advantages, PFGs have become the default mode for NMR data acquisition and are even used automate magnetic field "shimming" in the process NMR instrument optimization.

ADVANCES IN PROBE TECHNOLOGY

Magic Angle Spinning (MAS)–When NMR spectra are acquired for solids using standard methods, extremely broad peaks are produced caused by high proton dipolar and chemical shift anisotropy. These terms vanish when NMR spectra are acquired for liquids and only the spin Hamiltonians for chemical shift and spin coupling contribute. The problem of line broadening in the NMR spectra of solids was largely overcome by the introduction of a technique known as magic angle spinning (MAS).[4] MAS requires a specialized NMR probe that spins the sample at a high rate (>5 kHz) while positioned at an angle of 54º44′ relative to the applied magnetic field.

Microprobe Design–Conventional NMR probes use glass tubes (5 mm O.D.) for sample introduction. While various sample volumes may be used, the active volume (interacting with the rf-coil) represents a volume of 220 µL. Because NMR is a concentration-sensitive detector, increased sensitivity may be achieved by dissolving the sample in a reduced solvent volume and/or using a microprobe designed to maximize the sample presented to the rf-coil. Microprobes built for use with smaller tubes (1.0–3.0 mm O.D.) were introduced in the early 1990s for this purpose and may be used with volumes of < 100 µL in some cases.[5] It is important to note that for both conventional and microprobe designs, the solution in the tube must extend beyond the area eclipsed by the rf-coils as not to introduce sources of inhomogeneity. More recently, nanoprobes designs have been introduced (40 µL) in which the entire sample volume is presented to the rf-coil.[6] Because of the effects introduced at the liquid-air interface, nanoprobes must be used with MAS.

Flow Probes (LC-NMR)–An entirely new field of NMR spectroscopy has grown up around the use of NMR probes which incorporate a flow-through design.[7] Flow probes have a detector volume of 100–200 µL and permit sample introduction via flow injection analysis (FIA) or liquid chromatography (LC-NMR). FIA has proven useful for automating sample introduction for use with 96-well technology. LC-NMR offers the advantage of direct mixture analysis, avoiding the need for sample isolation. Although LC-NMR has been used with a number of applications,[8,9] a fundamental limitation of the technology results from the disparate time frames associated with HPLC and NMR data acquisition.

Very recently, a microflow NMR probe has been introduced having a volume of only 5 µl (2 µl active RF coil volume).[10] This design, constructed largely from fused silica capillary tubing, allows for the introduction of highly concentrated samples, reduces the consumption of deuterated solvents, and minimizes post-column band broadening during LC-NMR.

Cryoprobes–One of the most effective ways found to reduce the noise in NMR spectra is to reduce the temperature of the probe. Superconducting probes are now commercially available which use low temperature rf coils to reduce the overall noise level. In these experiments the coils are typically maintained at 20–25ºK while the sample is left at room temperature. In most instances, a 4-fold reduction in noise can be expected using current technology.[11] It is important to note that an increase in signal-to-noise of 4-fold reduces the time needed for NMR acquisition by more than a factor of ten.

Magnet Size–Throughout the history of FT-NMR, superconducting magnets have steadily increased in size. At the time of this writing, the largest commercially available magnets are on the order of 900 MHz. In addition to providing greater spectral

resolution, high field magnets allow data acquisition to proceed more rapidly since less time is required to achieve a usable resolving power. Despite the performance of such magnets, ultra high field instruments are not widely deployed in the field owing to their high cost. A more common magnet size used in the pharmaceutical industry today is 600 MHz.

SPECTROSCOPY

Continual advances in instrumental methods of analysis have helped to establish these techniques as the mainstream of the analytical laboratory. The conventional wet chemical methods are gradually playing a minor role in the analytical discipline. S*pectrometric methods,* instruments based on the absorption or emission of electromagnetic (EM) radiation as a result of its interaction with matter are described and their applications are explored. These include ultraviolet (UV), visible, infrared (IR), florescence, Raman and light scattering techniques.

A study of the theory and applications of spectrometric methods of analysis necessitates a brief understanding of electromagnetic (EM) theory. Maxwell first expressed the concept of the electromagnetic field in 1860. His equations theorized the existence of waves that travel through electromagnetic fields and whose properties are identical to those of light. The oscillation of an electron gives rise to EM radiation. As is illustrated in Figure 34-17, at each point in the direction of the beam, the electric field and magnetic field, represented by two vectors, are perpendicular to each other. The wavelength, λ, is defined as the distance between successive maxima or minima, and is expressed in nanometers (nm) or 10^{-9} meters, formerly known as Angstroms (Å), (one Å = 10^{-8} cm). The frequency in cycles per second (cps or Hz) is denoted by ν. The frequency is related to λ by $\nu = c/\lambda$, where c is the velocity of light in vacuum. The time required for the completion of one cycle is designated by τ, which is related to ν by $\tau = 1/\nu$. The reciprocal of wavelength, $1/\lambda$, is referred to as wave number, $\underline{\nu}$, expressed in reciprocal centimeters, cm^{-1}. The wave number is employed particularly in describing the position of peak maxima for IR spectra.

Planck, in 1900, formulated a concept of quantum restriction. He stated that oscillating atoms of a hot body can have only energies that are integral multiples of $h\nu$. In other words, the energy of an oscillator is discontinuous and any change in the energy can occur only by a jump between two energy states. Planck showed that the energy in a photon of light is related to wave frequency by the expression $E = h\nu = hc/\lambda$, where h is Planck's constant, 6.6256×10^{-27} ergs/sec. In 1903 Einstein conducted his experiments on the photoelectric effect of light. He concluded that electrons are emitted from the surface of a specific metal upon its illumination with light of a relatively low wavelength such as blue light.

Red light, irrespective of its intensity, fails to eject an electron from a similar metal. These findings by Michelson and Morley, Planck, Einstein, and others could not be explained by Maxwell's assigned wave properties. Considering these facts, a reliance on the dual nature of light, behaving both like a wave and a particle, seemed to be indispensable for resolving many physicochemical phenomena.
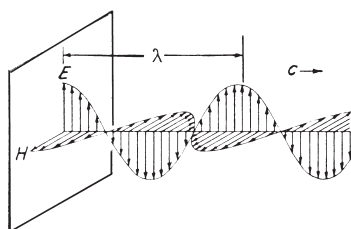
Figure 34-17. A plane-polarized electromagnetic radiation, *E,* electric vector, *H,* magnetic vector.

Molecular Interactions and EM Radiation

The presence of radiation of a particular frequency is necessary, but is not always sufficient, to induce a change in the energy level of a molecule. Quantum restrictions specify certain conditions for the interaction of radiation with a molecule. On many occasions energy is absorbed only if the radiation frequency corresponds to the components of the molecular frequency. This is referred to as resonance absorption.

The position of maximum absorption, λ_{max}, for a molecule in a particular region of the spectrum is a function of the total structure of the molecule with a transition energy corresponding to a given wavelength. The intensity of the absorption maximum, ϵ_{max}, is a function of the probability of EM radiation- molecule interaction and polarity of the excited state. At ground state (i.e. room temperature) a molecule is normally in its lowest energy state. The transition between E_1 and E_2, two energy states or levels of a molecule, occurs by the interaction of EM radiation with a molecule. The difference between E_1 and E_2 is designated by ΔE, whose frequency of radiation is expressed as $\Delta E = h\nu$ ergs.

Very high energies ($>10^8$ cm^{-1}) disturb and cause changes in the nucleus of the atom regardless of its environment. Lower energy however, causes a change in the electronic distribution around the nucleus.

Regions of the Spectrum

The whole range of EM radiations are divided arbitrarily into a number of regions. Interaction between a molecule and various kinds of EM radiation gives rise to a change in the electronic energy and/or kinetic energy of the molecule. In most cases, the energy absorbed is converted quickly to vibrational, rotational, and translational energy. However, in specific cases, emission occurs either immediately as in *fluorescence,* or after a short time as in *phosphorescence.* These specific changes in the energy of a molecule result in the generation of a characteristic spectrum that can be used for both structural elucidation and quantitative determination. Figure 34-18 depicts a wavelength and frequency scale for the different regions of the EM radiation spectrum.

A theoretical and practical description of various types of spectrometry of primary interest in the pharmaceutical industry is given in the following sections. The length of discussion of each topic is based on the extent of the applicability of the method in pharmaceutical analysis.

Absorption Spectrometry

Absorption spectrometry is the measurement of the selective absorption by atoms, molecules, or ions of electromagnetic radiation having a definite and narrow wavelength range, approximating monochromatic energy. Absorption spectrometry encompasses the wavelength regions; ultraviolet (200–380 nm), visible (380–780 nm), near infrared (780 nm to 2.5 μm) and infrared (2.5–40 μm). The region between 10 nm and 200 nm, known as the far UV or vacuum UV (as it requires the complete absence of air due to its interference), has minimal application in pharmaceutical analysis. Atomic absorption spectrometry involves the measurement of radiation absorbed by the unexcited atoms of a chemical substance that has been aspirated into a flame or other high-energy sources.

Theory

When electromagnetic radiation travels through a medium containing atoms, molecules, or ions, a number of events may take place.

- The intensity of the emergent energy may be identical to the intensity of the incident energy. This indicates that no absorption of radiation has occurred.
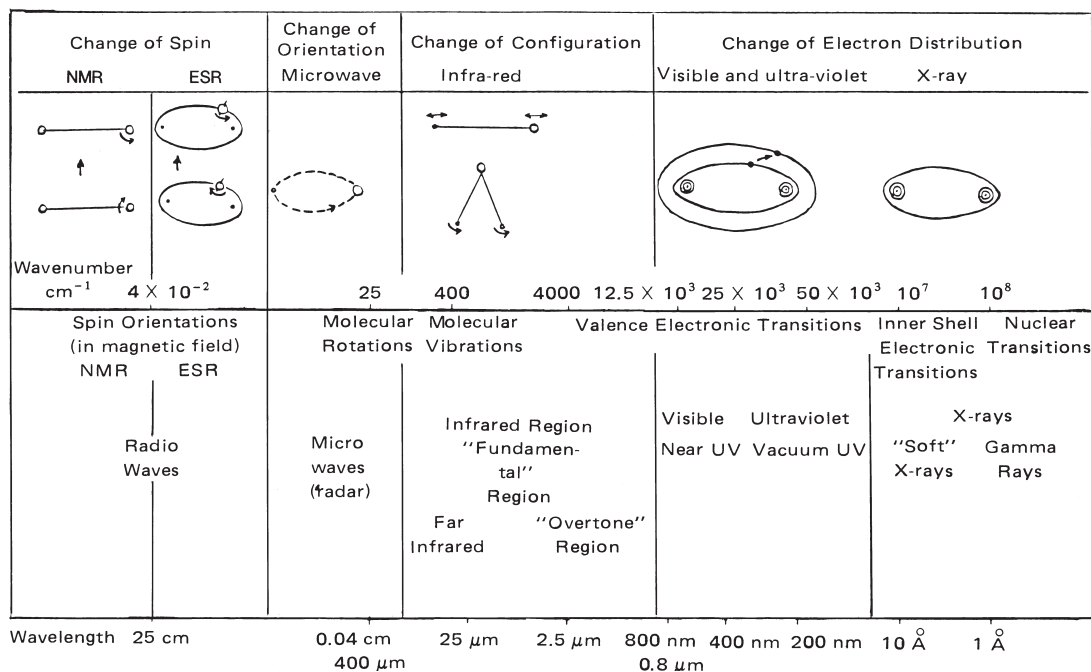
Figure 34-18. The electromagnetic spectrum.

- Reflection, refraction, and/or scattering may occur.
- The intensity of the emergent energy is less than that of the incident energy.

This latter condition indicates that some absorption has taken place (absorption spectrometry). As a result of this absorption, the species involved are activated from their lowest energy state (ground state) to higher energy states (excited states). For absorption to occur, the energy of the exciting radiation must match the quantified energy difference between the ground state and one of the excited states of the species. In atomic absorption, excitation occurs only through electronic transition. In visible and ultraviolet spectrometry, radiation energy can excite only the outermost or valence electrons. Accompanying the electronic excitation (E_e) is a change in vibrational energy (E_v) and rotational energy (E_r) of the molecule. For polyatomic molecules, vibrational and rotational transitions can occur in addition to electronic excitation. As a result, the molecular spectrum consists of closely spaced absorption bands instead of the sharp lines as in atomic absorption. Pure vibrational and some rotational transitions can be achieved by infrared radiation.

The duration of the excited state is brief (10^{-8} to 10^{-9} sec), its existence being terminated by any of several *relaxation* processes. The most common relaxation occurs with the production of heat, which may cause a slight increase in the temperature of the medium. Another form of relaxation results as the decomposition of the excited state into new species (photochemical reactions) according to

$$M + h\nu \rightarrow M^* \text{ (excited state)}$$

$$M^* \rightarrow M + \text{heat}$$

$$M^* \rightarrow M' \text{ (new species)}$$
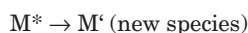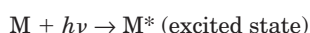
Alternatively, relaxation may result in emission of radiation at specific wavelengths characteristic of the excited species (emission spectroscopy), or in emission of radiation at longer wavelengths than the incident beam, immediately (fluorescence) or after a short time (phosphorescence).

Ultraviolet and Visible Absorption Spectrometry

The UV and visible absorption bands are due to electronic transitions in the region of 200 nm to 780 nm. In case of organic molecules, the electronic transitions could be ascribed to σ, π, or n electron transition from the ground state to an excited state (σ^*, π^*, or n^*). Because the σ electron is involved firmly in the construction of a single bond, its transition requires much more energy (usually in far UV) than the n electron (nonbonding electrons) or less tightly bonded π electrons.

There are four types of absorption bands that occur due to the electronic transition of a molecule:

R-Bands: $n \rightarrow \pi^*$, in compounds with C=O or NO_2 groups $\epsilon_{max} < 100$

K-Bands: $\pi \rightarrow \pi^*$, in conjugated systems $\epsilon_{max} > 10,000$

B-Bands (benzenoid bands): due to aromatic and heteroaromatic systems, $\epsilon_{max} < 2000$

E-Bands (ethylenic bands): in aromatic systems, ϵ_{max} 2000 to 14,000

Beer's Law

If incident light with wavelength λ and intensity I_0 impinges on a solution with concentration c, and pathlength l of 1 cm, the radiant energy of the light decreases in an exponential fashion. Thus, if a given concentration of a substance absorbs 50% of the incident radiation, doubling the concentration will not absorb 100% but rather 75% of the light. The thickness of the sample or pathlength has a similar effect on the absorption. Mathematically, the radiation-concentration and radiation-pathlength relation can be expressed by

$$\frac{dI}{dc} = -k_1 I \text{ and } \frac{dI}{dl} = -k_2 I \qquad (6)$$

Figure 34-19. The UV absorption spectrum of salicylic acid.

Figure 34-20. A representative Beer's law plot.

The wavelength corresponding to maximum absorptivity, ϵ_{max}, is denoted by λ_{max}.

A few of the most generally employed terms in absorption spectrometry are given in Table 34-5.

Quantitative Applications of UV and Visible Spectrophotometry

One of the major uses of UV and visible spectrometry is for quantitative analysis. An unknown concentration of a known compound, if it conforms to Beer's Law, can be determined by using Equation 8. A representative calibration curve, shown in Figure 34-20 is constructed by plotting absorbance (A) versus concentration.

For quantitative estimation, the samples for UV absorption can be examined in the form of a vapor or a solution. Both polar and nonpolar solvents can be employed to prepare an analytical sample. The cutoff point of a solvent, however, should be recognized as it renders the solvent useless at wavelengths below this value. This is the wavelength at which the absorbance of a solvent approaches unity, using water as a reference. The cutoff points for many solvents can be found in the literature and in solvent charts supplied by several suppliers of solvents.

An understanding of the limitations of Beer's law must be taken into consideration. Some of these are of such a fundamental nature that they constitute a real limitation of the law. The Beer's law during quantitative analysis does not take into consideration the effects of pH, temperature, wavelength, or solute–solvent and solute–solute interactions, such as association (intermolecular hydrogen bonding), dissociation, and chemical reaction. Because of these limitations, the law usually applies only to dilute solutions, where these interactions are insignificant. Another limitation to the Beer's law is the inability of most instruments to provide monochromatic radiation.

A simplified diagram of a UV-visible spectrometer is presented in Figure 34-21 and its major components are outlined in the Table 34-6.

Integration of Equation 6 gives:

$$\int_I^{I_0} \frac{dI}{I} = -k_1 \int_0^c dc \text{ and } \int_I^{I_0} \frac{dI}{I} = -k_2 \int_0^l dl \qquad (7)$$

Evaluation of the integrals between limits, combining the two formulas, and incorporating the value 2.303 (for transforming the natural log into a log of base 10) in the constant provides the more familiar equation used in spectrometry,

$$\log (I_0/I) = \epsilon c l \qquad (8)$$

Where I_0 is the intensity of the incident energy, I is the intensity of the emergent energy, c is the concentration, l is the thickness of the medium (in cm), and ϵ is the molar absorptivity (formerly expressed as molar extinction coefficient) for concentration in mol/L.

If the concentration is expressed in g/L, absorptivity is designated by a instead of ϵ. The term $\log I_0/I$ or $\log (1/T)$ is referred to as absorbance, A (formerly stated as optical density or extinction); T is Transmittance or I/I_0. $E_{1cm}^{1\%}$, which is encountered less frequently in the literature, represents a concentration of 1% w/v and a 1-cm cell thickness and is used primarily in the investigation of those substances of unknown or undetermined molecular weight (usually impure natural products).

A typical UV absorption spectrum, shown in Figure 34-19, is the result of plotting wavelength (λ) versus absorptivity ($\log \epsilon$).

Table 34-5. UV Terminologies

Chromophore	A moiety of molecule responsible for selective absorption of radiation in a given range
Auxochrome	A chemical group which does not give rise to an absorption band by itself, but upon being attached to a chromophore alters both the position and/or intensity of the peak
Bathochromic shift	A shift of the peak position (λmax) to a higher wavelength due to the effect of a substituent or solvent (red shift)
Hypsochromic shift	A shift in (λmax) to lower wavelength (blue shift)
Hyperchromic and Hypochromic shift	An increase and decrease in absorptivity

Figure 34-21. A classic UV-visible spectrometer (newer instruments use a grating instead of prisms).

Table 34-6. Comparative Characteristics of Spectroscopic Methods of Analysis

TECHNIQUE	WAVELENGTH (nm)	SOURCE	DETECTOR	SAMPLE	INFORMATION TYPE	APPLICATION
Absorption Spectrometry Ultraviolet/Far UV	200 to 380 nm	Hydrogen or deuterium lamp	Photomultiplier tube (photodiode array, Photon diode) and semiconductors (Charge transfer devices)	Vapor/solution	Little structural information but the presence of unsaturated sites in the molecule	Qualitative and quantitative analysis, confirmation analysis, multicomponent analysis, derivative spectrometry
Visible	380 to 780 nm	Tungsten lamp or deuterium arc lamp	Photomultiplier tube and semiconductors	Vapor/solution	Presence of unsaturated sites in the molecule	Quantitative analysis, confirmation analysis (purity control)
Infrared	2.5 to 40 μm	Nernst or globar unit	Thermocouple or bolometer	Gas, liquid or solid (NaCl, KBr, and CsBr pellets)	R-H vibrational mode	Characterization of molecules
Fourier transform Infrared (FT-IR)	2.5 to 40 μm	Zirconium oxide or rare earth oxides (Nernst Source), silicon carbide	Mercury cadmium telluride (MCT), deuterated triglycerine sulphate (DTGS) crystal or lithium tantalite (LiTaO$_3$)	Same as IR	Structural analysis	Qualitative powers of FT-IR coupled with separation technique as GC-FT-IR and LC-FT-IR
Diffuse Reflectance (Specular or diffuse or attenuated total reflection)	2 to 10 μm	Same as IR	Same as IR	Sample is diluted with KBr powder	Structural analysis	IR spectra of solid samples ie, drugs, pharmaceuticals, food products, soap powder, coal, clay, paper, painted surfaces, polymer foam, catalysts.
Infrared Microscopy	Same as IR	Same as IR	Same as IR	Same as IR	Same as IR	Flaws and variations in bulk properties of matrices.
Pattern Recognition Analysis	UV and IR regions					Identification and differentiation of plastic materials used in Pharmaceutical packing
Hierarchical Cluster Analysis	UV and IR regions					
Emission Spectrometry		AC-, DC, and AC spark				Qualitative detection of all metals and nonmetallic elements
Flame Photometry		No light source			Qualitative and quantitative analysis	Group IA and IIA metals (Quality control measurement of alkaline or alkaline earth metals)
Plasma Emission		No light source/ hollow cathode lamp				Elemental Analysis
Atomic absorption spectrometry		Discharge lamps (argon or neon)		Solution or in solid state	Qualitative/ quantitative analysis	
Fluorescence Spectrometry	Visible or UV range	Xenon arc lamp				PAH analysis in water, measurement of aflatoxins
Raman Spectrometry	4000 to 25 cm^{-1}	Helium/neon laser	Photomultiplier detector		Vibrational and rotational energy modification	Qualitative/quantitative analysis of inorganic, organic, and biological systems

In general, in single beam spectrometer the sample is placed in the compartment where it encounters monochromatic energy. In a double-beam instrument, this compartment contains a beam-chopping device or a beam-switching assembly that allows the beam to pass alternatively through the sample and reference cells (about 35 times/sec). This allows the sample-reference relationship to remain unaffected by slight changes in the source or optics of the instrument. The detector is usually a photomultiplier tube. The output from the detector is amplified and observed on a meter, a recorder or a cathode ray tube. Most next generation spectrometers are equipped for automatic and continuous recordings. Spectrometers employing the latest technology can be interfaced with a digital computer through an analog to digital converter for the direct determination of difference spectra of analytes as well as for the storage of reference spectra.

Modern Spectrometric Techniques

There has been significant progress in the use of holographic gratings and microprocessor control in the design of modern spectrometers. Recent models feature automatic control of all operating parameters such as wavelength selection and calibration, baseline correction, programmed scanning, first-, second-, third-, and fourth-derivative spectra, light-emitting diode (LED) readouts of absorbance or concentration in addition to screen monitoring and hard-copy printouts.

Also, due to the advent of stable microelectronics, in addition to the wide availability of microprocessor controlled and fully automated spectrometers, new interest is rising in all of the UV-visible absorption techniques that normally require substantial instrument control and data manipulation. These include simultaneous multicomponent analysis, reaction rate determinations, and dual-wavelength derivatives. Also, there is a significant increase in the use of *Difference Spectrometry* as a means for increasing sensitivity, improving detection limits, and decreasing noise as compared to conventional absorption spectrometry.

Other new high-sensitivity spectrometric techniques have gained wide attention recently, especially in trace analysis application and determination of solvent spectra. These new techniques include

- Laser-absorption intracavity techniques based on the dyelaser oscillating mechanism, which are capable of measuring absorbance's in the 5×10^{-6} range and are more suitable for aqueous systems.
- Wavelength modulation (peak-sensing) methods suitable for measuring two different samples simultaneously as well as double derivation of reflectivity, with a sensitivity of up to 1×10^{-7} g.
- Colorimetric methods for the measurement of energy absorbed by the solution using laser sources. These methods have a range of detection between 1×10^{-7} to 8×10^{-8} g and they include thermocouple colorimetry, photoacoustic colorimetry, and thermal lens techniques. Both photoacoustic and thermal lens methods suffer great loss of sensitivity in aqueous media.
- Photon-counting and diode-array detectors.

Diode-Array UV-Visible Spectrophotometry

Advances in technology have led to the development and implementation of photodiode detectors, which when placed in closely spaced linear arrays, offers rapid and accurate spectrum analysis. The primary advantage of linear-array detectors is that they permit the simultaneous analysis of an entire spectrum over a period of a few seconds. This is advantageous when performing kinetic studies involving rapidly changing events. A simplified diagram is presented in Figure 34-22.[12]

Diode-array detectors have an added advantage of increased wavelength resolution. Precision-matching of slit sizes to individual photodiodes and focusing the spectrum in a focal plane can enhance wavelength-resolving power to 1 to 2 nm.
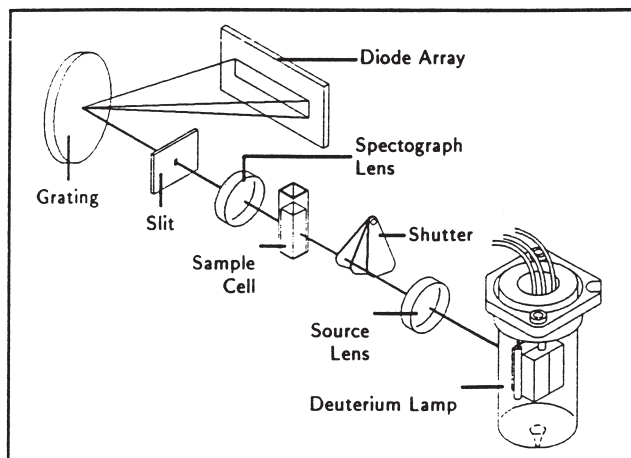
Figure 34-22. HP 8452A optical system with deuterium lamp.

Infrared Spectrometry (IR)

The range of EM radiation between 0.8 and 500 μm is referred to as infrared radiation. At present the IR spectrometer is one of the instruments most frequently employed in the characterization of organic molecules. Unlike the UV-visible spectral plots, the IR spectrum usually is represented with percent transmittance, rather than absorbance, as the ordinate (X-axis). Also, it is customary to use the unit of reciprocal centimeter (cm^{-1}) or the wave number for the abscissa (Y-axis) rather than the wavelength. This is because of the direct proportionality between the wave number and the energy as well as the frequency of the radiation; the frequency can, in turn, be related directly to molecular vibrational frequencies. An example of an IR spectrum is shown in Figure 34-23.[13] The most commonly used region of the IR spectrum in pharmaceutical chemistry is the region between 2.5 μm $(4000 \ cm^{-1})$ and 16 μm $(625 \ cm^{-1})$.

The near infrared region (NIR) or the overtone region refers to the segment from about 700 nm $(12,500 \ cm^{-1})$ to 2.5 μm $(4000 \ cm^{-1})$; the far infrared region (FIR) or the rotational region is between 400 and 20 cm^{-1}.

Theory

In order for IR radiation to be absorbed by a molecule, two criteria must be met: the molecule should possess a vibrational or rotational frequency identical to that of the impinging EM radiation, and a net change in the magnitude or direction of the dipole moment should occur as a result of radiation-molecule
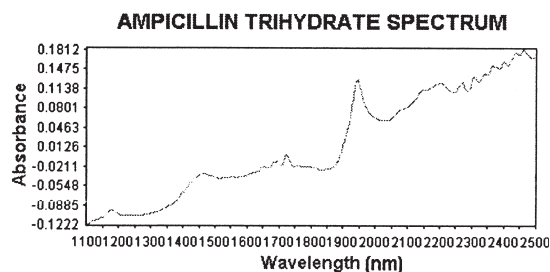
Figure 34-23. NIR diffuse reflectance spectrum of ampicillin trihydrate.

interaction. When IR radiation impinges upon a molecule at the suitable frequency, the vibration and/or rotation of the molecule is altered. If the frequency of the impinging EM radiation matches a natural vibrational frequency of the molecule, a net transfer of energy occurs that creates greater amplitude of vibration and, as a result, absorption of radiation occurs.

The longest wavelength (lowest energy) of IR radiation that induces a change in the vibratory motion of a molecule gives rise to an absorption band known as the *fundamental band.* There is only one fundamental band in a diatomic molecule, although multiples of the band frequency (ν), known as overtones, can occur as 2ν, 3ν, and so on.

Rotation of asymmetric molecules around their centers of mass results in a periodic dipole change that interacts with the incident EM radiation causing a higher frequency of the molecular rotation and absorption of radiation occurs. The energy required to cause a change in rotational levels only is very small (100 cm^{-1}) and comprises the far IR (FIR) region. Absorption by gases in this region appears as discrete, well-defined lines. However, because of the intramolecular collisions and interactions in liquids and solids, broadening of the absorption lines occurs and usually appears as a continuum. The FIR region, which experimentally is difficult to study, has limited application in pharmaceutical chemistry and thus will not be discussed further.

Because absorption of IR radiation alters both vibrational and rotational characteristics of a molecule, absorption bands are not defined lines but are bands that are centered upon one frequency. As the total kinetic energy is a combination of translational, rotational, and vibrational energies of a molecule (ie, $E_t = E_{tr} + E_r + E_v$), a polyatomic molecule consisting of n atoms will have 3n degrees of freedom of motion. The possible fundamental vibrational modes of a molecule can be calculated by subtracting 3 for translational energy and 3 for rotational energy (2, if the molecule is linear). This gives a total of 3n-6 possible vibrational modes. The theoretical number of fundamental absorption bands, however, is not observed due to such factors as weak absorptivity, coalescence of several closely located bands, and lack of required change in dipole moment. Because the 2 to 16 μm region normally employed for IR investigation covers both fundamental and *overtone* regions, the total number of absorption bands in an IR spectrum may greatly exceed the theoretical number.

The atomic stretching vibration can be approximated mechanically by Hooke's law, $F = -kx$, where F is the restoring force, k is the proportionality or the force constant (dyne/cm), and x is the displacement distance. For a diatomic molecule with atoms of masses m_1 and m_2, the frequency of fundamental vibration is expressed by

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \qquad (9)$$

or in terms of wave number by

$$\bar{\nu} = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}} \qquad (10)$$

Where μ is known as the reduced mass, defined by

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \qquad (11)$$

Application of the equation for the C—H stretching frequency with $k = 5 \times 10^5$ dynes/cm, $m_1 = 19.8 \times 10^{-24}$ g, and $m_2 = 1.64 \times 10^{-24}$ g gives the value 3040 cm^{-1} (slightly higher than the observed value, 2950 cm^{-1}, which is caused by neglect of the environmental effect). The vibrational modes of a CH_2 group are depicted in Figure 34-24. It should be observed that more energy is required for the stretching vibration than for the bending vibration.
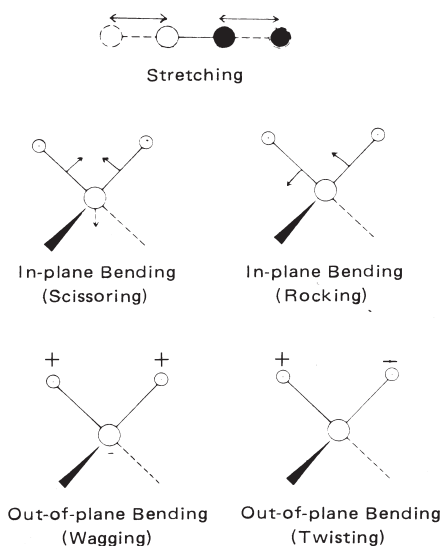
Figure 34-24. Types of molecular vibrations. Plus sign (+) indicates motion from plane of page toward reader, Minus sign (−) indicates motion from plane of page away from reader.[15]

The position of the absorption bands is determined by the symmetry of a molecule, the masses of atoms, the force constants of the chemical bonds, and the interaction of vibrations (Fermi interactions). Hydrogen bonding affects the position of the bands by shifting the frequency of the stretching vibration to a lower frequency and that of the bending vibration to a higher frequency.

Characterization of Molecules

There are two major applications of IR spectrometry in the characterization of various molecules: determination of the identity of a compound by means of spectral comparison with that of an authentic sample, and verification of the presence of functional groups in an unknown molecule. The latter aspect is quite important in the structural elucidation of synthetic organic compounds or substances isolated from natural sources.

The position of the absorption bands due to stretching and in-plane bending vibrations of the functional groups, such as C=O, C—H, N—H, O—H, are somewhat independent of the influence of the neighboring groups in the molecule. These bands usually occur at 4000 to 1300 cm^{-1}. The position of the bands below 1300 cm^{-1} is influenced markedly by neighboring groups in the molecule. The portion of the spectrum from 1300 to 400 cm^{-1} is referred to as the "fingerprint" region.

Extensive charts and tables of the characteristic group absorption frequencies for common organic functional groups can be found in many of the texts listed in the *Bibliography*. Several catalogs of reference spectra have been published, the most voluminous of which is that of the Sadtler Research Laboratories, currently in excess of 90,000 spectra.

The C—H stretching and bending vibrations occur at 3300 to 2800 cm^{-1}. Each type of hydrocarbon has its own characteristic band position; for example, saturated acyclic and cyclic hydrocarbons have stretching ν at 2960 to 2850 cm^{-1} and in-plane bending ν at 1470 to 1360 cm^{-1}; unsaturated olefinic C—H stretching at 3090 to 3000 cm^{-1} and unsaturated acetylenic C—H stretching ν at 3300 to 3270 cm^{-1}; aromatic C—H stretching ν at 3100 to 3000 cm^{-1} and the out-of-plane bending is at 900 to 650 cm^{-1}. The most characteristic band for aromatic compounds, however, is at 1610–1590 cm^{-1} (due to aromatic skeletal vibration). The characteristics of band frequencies are given in Table 34-7.

Table 34-7. Characteristics of Band Frequencies

O—H Vibration	Stretching at 3700 to 3350 cm^{-1}, depending on the extent of hydrogen bonding.
C—O Vibration	Stretching at 1280 to 1000 cm^{-1}, depending on whether it is an alcohol, phenol, ester, ether, etc.
C=O Vibration	Stretching at 1950 to 1640 cm^{-1}, These bands are quite intense and very conspicuous. Hydrogen bondings, field effect, and conjugation affect the position
N—H Vibration	Stretching at 3500 to 3300 cm^{-1}, hydrogen bonding at lower frequency. Bands for N$^+$H$_3$, N$^+$H$_2$, and N$^+$H occur at about 3200, 2700, and 2000 cm^{-1}, respectively.
C—N Vibration	Stretching of aliphatic compounds at 1210, aromatic at 1250 to 1350, C=N at 1680 to 1640, and for C≡N at 2250 cm^{-1}

Quantitative IR

IR spectrometry generally is employed for qualitative identification, with limited use in quantitative analysis. Because of the uniqueness of IR spectra, quantitative methods may not require prior separation of the analyte from excipients. The sensitivity of IR analysis, however, is poor, only 0.01 to 0.001 of the sensitivity of UV, and, therefore, it has only a few applications in quantitative analysis. The major components of IR spectrometer are illustrated in Figure 34-25. The most commonly used prism materials for dispersion of IR radiation are: 1) NaCl with a refractive index of 1.5442. This provides good dispersion at 2000 to 650 cm^{-1}, but poor dispersion beyond 2000 cm^{-1}; 2) KBr, with a refractive index of 1.53, disperses at 1600 to 370 cm^{-1}; and 3) CsBr, with a refractive index of 1.69, disperses at 1000 to 250 cm^{-1}. In recent years grating systems have been employed widely than the prism, primarily because of their high resolving power.

As seen in Figure 34-25, the source beam is reflected by mirrors to form the sample and reference beam. After passing through the sample and reference, the beams are chopped by a mirror that serves to focus each beam alternately onto the entrance slit of the monochromator. If the sample absorbs part of the radiation, the intensity of the two beams will be unequal.

This inequality results in the development of an out of balance signal in the detector. After amplification and rectification, the signal is relayed to a comb or wedge to drive the reference beam attenuator to reduce the intensity of the reference beam. As the difference between the two beams becomes zero,

Figure 34-25. The optical system of a classic IR spectrometer (newer instruments use gratings instead of prisms).

the out of balance signal also becomes zero. The pen of a recorder, which is connected to the attenuator, will perform the function of plotting the absorption coordinates on a paper chart. The abscissa of the chart is a function of frequency, and the resulting tracing of percent transmission versus frequency is known as an IR spectrum.

Fourier Transform Infrared Spectrometry (FT-IR)

The wide availability of high-powered microcomputers at reasonable cost has helped popularize the applications of transform spectroscopy in general and Fourier transform in particular to several branches of spectrometry. These include IR, NMR, and MS. FT-IR, however, has been one of the first techniques developed; today, it is the instrumentation of preference over dispersive IR for handling ever smaller and more complex samples. Superior sensitivity and resolution, absolute wavelength accuracy, and higher precision of measurements are some of the reasons behind the rapid growth of FT-IR.

Basically, the technique is a coupling of a Michelson interferometer with a sensitive infrared detector. However, because of the enormous amount of data generated, a microcomputer is essential for data handling. In the Michelson interferometer, there is no monochromator and radiation of many frequencies passes through the sample. The source radiation is split between a fixed mirror and a movable one. The two reflected beams then are combined, either constructively or destructively, at the beam splitter, depending on the position of the movable mirror. As the path difference between the two beams is altered and, because only the nonabsorbed frequencies reach the detector, the signal pattern becomes the sample interferogram. For monochromatic radiation, the amplitude of the signal is a cosine function of the mirror position. For polychromatic radiation, the signal is a summation of all the constructive reinforcement or destructive interferences of each wavelength interacting with every other wavelength and results in a unique interferogram for each particular sample.

To handle the complex mathematical treatment needed for calculations, it was found that the cosine Fourier transform can relate the intensity of the interferogram as a function of the mirror travel, $I(x)$ (Eq 12) and the intensity of the frequency $I(\nu)$ (Eq 13) of the IR radiation:

$$I(x) = \int_{-\infty}^{\infty} I(\nu) \cos (2\pi\nu x) \, d\nu \qquad (12)$$

and after calculating (using a computer) the inverse transforms,

$$I(\nu) = \int_{-\infty}^{\infty} I(x) \cos(2\pi\nu x) \, dx \qquad (13)$$

by which the interferogram could be related back to the IR spectrum.

Modern FT-IR spectrometers provide full spectra that can be monitored continuously on a CRT screen while scanning. Standard software packages include spectral subtraction, baseline correction, integration, peak selection, multicomponent and factor analysis, quantitative analysis, and spectral library searching. The use of a new mercury cadmium telluride (MCT) detector, diffuse reflectance accessory, cylindrical interval reflection device, and transmission or reflectance microscopy are recent features that enhance the instrument's sensitivity and versatility.

Diffuse Reflectance

The diffuse reflectance technique has become very popular in recent years and is applicable to a wide range of solid samples. In this technique, IR radiation is focused on a sample and the
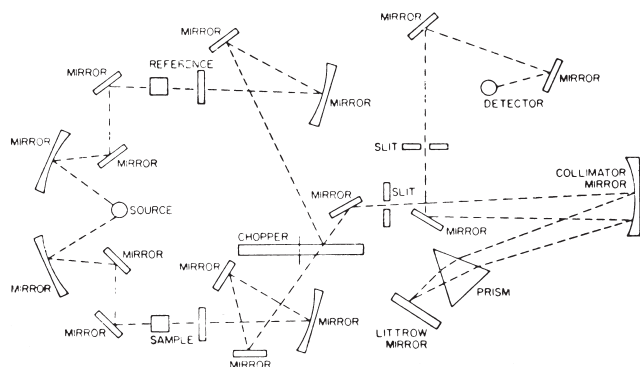
reflected or scattered radiation is collected over a wide solid angle, hence the term *diffuse reflectance*. Sample handling is simple and straightforward. Extensive grinding or use of high pressure is not needed, which eliminates the risk of altering the sample structure. Some of the samples for which good IR spectra are obtained are drugs, pharmaceuticals, food products, soap powder, coal, clays, paper, painted surfaces, polymer foam, and catalysts. Many samples, such as inorganics, may be run in neat form. Dilution with KBr powder is often used to reduce the intensity of strong absorption bands.

Infrared Microscopy

Often the analyst is confronted with a sample that precludes exhaustive sample handling. At times these samples take the form of minute flaws or variations from the bulk properties of the matrix. Such samples arise from many sources: forensic, textile, packaging, polymers, films, coatings, paper, and electronic components, all of which can contain small discontinuous areas of questionable composition. When these imperfections arise, the analyst is asked to examine the spot or fragment so that its source in relation to the manufacturing process can be identified.

Infrared spectrometry is especially useful in these situations because it is nondestructive. In addition, the variety of sample-handling techniques available often permits a sample to be analyzed *in situ*. For those problems where it is impossible to extract the critical portion of the sample, this is especially important. Microsampling techniques are employed in IR to improve sensitivity or to restrict the field of view, and thus, eliminate gross background interference. Beam condensers and other magnification tools have been used for many years. Microscopes extend the utility of IR to samples of the order of 10 µm. Use of a microscope allows alignment of the small sample in the IR beam as well as focusing the energy.

Thermogravimetric analysis (TGA) coupled with FT-IR is a new way of monitoring the evolved gases generated during sample decomposition or volatilization caused by heating over time. Using TGA, identification of the components of the gases can be used to determine sample characteristics.

Advantages and Limitations of FT-IR

The speed and high sensitivity of FT-IR, which make it ideal for microanalysis, arise from two factors.

1. The use of what is known as the multiplex or the Fellgett advantage where a very high signal-to-noise ratio exists due to the fact that the sample; thus, the detector is affected by all frequencies at one time.
2. The radiation power throughput of the interferometer is significantly larger than for the dispersive instrument (about 40 times).

These advantages of FT-IR make it the technique of choice for coupling the qualitative power of IR to such separation techniques as gas and liquid chromatography (GC-FT-IR and LC-FT-IR).

Pattern Recognition Analysis

As the preceding sections of this chapter have shown, the spectra obtained in the UV and IR regions are very useful. This utility has been extended by the implementation of pattern recognition analysis. Subtle differences between data sets can be visualized easily, resulting in faster identifications and much quicker decisions. Two approaches, *Principal Component Analysis* (PCA, referred to previously in the near IR discussion), and *Hierarchical Cluster Analysis* (HCA) are mentioned here.

In PCA, new sets of variables or factors that are linear combinations of the original variables in the data set are calculated.

The few dimensions in the new factor space that are needed to represent all of the significant information in the data are called principal components, whereas other factors represent noise components only. This view allows visualization of the natural clustering in the data, identifies outliers, and facilitates assignment of chemical or physical meaning to the data patterns that emerge.

The primary purpose of HCA is to present data in a manner that emphasizes the natural groupings in that data set. Distances between the samples (variables) in a data set are calculated and compared. When distances between samples are relatively small, the implication is that samples are similar. HCA results are presented in the form of a dendrogram; a tree-shaped distance map constructed using the sets of intersample distances. Dendrograms show clustering, and the branch lengths are proportional to the distances between the connecting clusters.

One of the activities that may benefit from the capabilities of pattern recognition analysis is the identification and differentiation of plastic materials used in pharmaceutical packaging. For a discussion of this very useful approach to data treatment, see the text by Massart, Vandefinste, and Deming in the Bibliography.

Emission Spectrometry, Flame Photometry, and Atomic Absorption

The study of atomic spectra is probably the most basic scientific phenomenon that has captured the curiosity and the imagination of physicists, astronomers, and chemists for centuries. The information gained from the intensive and unrelenting pursuit to establish its fundamental theories was a major factor behind the development of our modern physical and chemical sciences.

Theory

When gaseous ions or an aerosol form of metals and some nonmetallic elements are heated to a high temperature, the kinetic energy of the atoms or molecules is increased. Collisions occurring at such an elevated energy incur a high probability of transforming the kinetic energy into excitation energy. The electronically excited species are unstable, and if no chemical reaction occurs after 10^{-4} to 10^{-7} sec, the energy is lost by emission of EM radiation in the UV and visible region, with wavelengths that are characteristic of the species under investigation.

Commonly employed methods of excitation are flame, ac arc, dc arc, and ac spark. Flame provides low-energy excitation and is used for easily activated substances; it has been used more recently with great success in Inductively Coupled Plasmas (ICP) where a high temperature argon torch is used to excite most atoms. Electrical excitation by discharge also is very effective in volatilizing and exciting samples and a temperature range of 4000 to 8000K is attainable by this method. An ac spark provides excitation energies greater than the arc and is produced by application of a high voltage (10–50 kV) across the electrodes. Excitation also can be achieved with an optical ruby laser. The optical system of a typical emission spectrograph is shown in Figure 34-26. A diffraction grating can be used in place of a prism for radiation dispersion.

The major application of emission spectrometry is in the qualitative detection of all metals and most of the nonmetallic elements. Detection limits lie in the ppm or ppb range. Quantitative application, which used to be limited, has grown very rapidly lately, especially with the introduction of inductively coupled plasma techniques and laser sources. Currently, emission spectrometry provides an excellent rapid technique for the simultaneous or sequential quantitative determination of up to 30 elements.
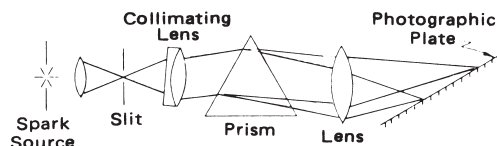
Figure 34-26. A simplified optical diagram of an emission spectrometer.

Flame Photometry (Flame Emission Spectrometry)

Flame photometry employs an emission-measuring device and uses a gas–air flame (1100°–1300°) for excitation. The detection is limited to group IA and IIA metals of the periodic table, which have a low-lying electronic level. Sodium is the most active in the series with a detection limit of 0.0002 ppm and beryllium is the least active with a detection limit of 25 ppm. The detection limit of a few elements is listed in Table 34-8.

Samples are dissolved in a solvent and introduced into the burner via an atomizer. Standard solutions used for analysis should be similar to the sample solution, as variables such as viscosity and temperature affect the nature of atomization, and thus the degree of excitation. In clinical laboratories the quantitative measurement of sodium, potassium, and calcium in biological samples is made by means of flame photometers.

Plasma Emission

Conventional atomization methods, such as combustion flame, furnaces, and electric arcs, are usually adequate for most of the traditional applications of atomic emission spectroscopy. These techniques, however, have several limitations, the most important of which are the instability of the atomization source, the possibility of chemical interaction such as metal oxide formation, the requirement of a relatively large size sample, low sensitivity, and finally, the inability to conduct simultaneous or sequential multielemental analyses. To overcome these limitations, new techniques called plasma emission spectroscopy have been developed.

Plasma is a partially ionized gas, usually a mixture of the sample vapor and a support gas. The plasma is generated electrically and once formed, a greater quantity of electric power can be transferred to it, raising its temperature to 9000K. Such a high temperature provides the analyst with a rich and stable source of atoms that act as a reservoir of free and highly excited atoms. The other advantages of plasma include a wide linear dynamic range, excellent sensitivity, high accuracy, and good precision. Also its suitability for simultaneous multielemental determinations at the ng/mL level has made it the method of choice for the analysis of trace constituents in samples of very limited volume.

Table 34-8. The Detection Limits of Some Elements Using Flame Photometry[a]

ELEMENT	WAVELENGTH (nm)	DETECTION LIMIT (ppm)
Barium	553.6	1.3000
Calcium	422.7	0.0030
Cesium	852.1	0.1000
Lithium	670.8	0.0020
Magnesium	285.2	0.2000
Potassium	766.5	0.0010
Sodium	589.3	0.0002

[a]Source: Courtesy Beckman.

Although there have been different types of plasma emission sources, the most popular sources that have gained wide application are the Direct-Current Argon Plasma and the Inductively Coupled Plasma.

Direct-Current Argon Plasma (DC Argon Plasma)

The main advantage of dc argon plasma is its excellent stability even in the presence of solvents, organics, and high acid or alkali concentrations. It usually consists of two carbon anodes, between which the plasma jet is formed, and a tungsten cathode. It requires about 1 kW of power and once ignited can be sustained by a low voltage. The plasma can sustain a temperature as high as 10,000K. Samples are introduced in an aerosol form, and their emission spectra are observed in a region isolated from the main plasma core, a procedure by which the sensitivity is enhanced greatly. Multielemental sequential analysis can be achieved easily with much lower detectable limits than with conventional flame emission.

The dc argon plasma also has a special advantage in the determination of trace amounts of arsenic and other nonmetallic elements. One limitation, however, is its unsuitability for automation, because the plasma supporting electrodes have to be replaced or reshaped after about 2 hours of operation.

Inductively Coupled Argon Plasma (ICP)

The main difference between inductively coupled argon plasma (ICP) and DC plasma is that the ICP derives its sustaining power by induction from a high frequency magnetic field. The pioneering work of Reed in the early 1960s laid the basis for ICP as an exciting new technique that can be used for the simultaneous determination of all of the periodic table elements with a lower limit of detectability in the ppb range.

ICP simply consists of a quartz tube (2.5 cm diameter) placed inside a coil that is connected to a high-frequency generator (4–50 MHz range) with output levels of 2.5 kW.[14] Because argon is a nonconductor, a seed of electrons (from a Tesla discharge coil) is first introduced before turning on the power. Argon is fed into the quartz tube and is ionized by the magnetic field produced by the induction coil. The seed electrons interact with the magnetic field and gain in intensity enough to ionize the gas flow and an eddy current, induced by the magnetic field, flows in circular closed paths around the discharge tube. After complete ionization, coned flame plasma is formed at the tip of the torch.

Because there is no electrode contact in ICP (as there is with dc plasma), the excitation and emission zones are separated from each other. This, besides the inert environment and the high temperature achieved, allows complete ionization of the sample with minimum chemical interference and a high signal-to-noise ratio of the sample's emission. These excellent conditions are the main reasons for the extreme sensitivity of ICP, typically in the ppb range.

Therefore, ICP offers the threefold potential of ultra-trace determinations on a multi-element basis, using a very small sample size (microliter or microgram level), in any type of matrix. ICP also is very amenable to complete automation and the simultaneous determinations of a vast array of both metals and metalloids.

Atomic Absorption Spectrometry (AA)

As early as 1860 Kirchhoff described the basic principles of atomic absorption (AA) spectra. It was not until 1955, however, that Walsh, Alkemade and Milatz demonstrated the theoretical background for its analytical applications. The simplicity of this technique makes it an attractive tool for the analysis of

many elements. At present, many chemical and clinical laboratories use this method for the quantitative determination of most of the elements in multivitamin and mineral formulations, drugs, and biological fluids.

Theory

In AA spectrometry, the elements are transformed into the atomic vapor form by drawing an aerosol of the sample solution into an open flame. A fraction, or most of the freed atoms, are then excited by exposure to a suitable source of radiation. The radiation absorbed by the unexcited atoms is related to the sample concentration. In this sense, AA then could be envisaged as the inverse of emission spectrometry, where the radiation emitted by the thermally excited atoms is related to concentration. It should be emphasized that usually the fraction of atoms excited by heat (via a flame or an electric arc) is relatively small for most elements. Also the atomic absorption of any element is generally at its resonance line—that is, a narrow range of wavelengths, usually in the UV or visible region of the spectrum, corresponding to the electronic transition between the lowest excited state and the ground state.

FACTORS AFFECTING AA SPECTRA

Solvents—In general, an organic solvent enhances the absorption signal, and therefore, it may alter the absorption intensity.

Anions—These can bond strongly with metals and tend to reduce the signal intensity. EDTA chelation could eliminate such effect.

Metal Binding—Sometimes, the presence of one metal interferes with the signal of another. For example, either Si or Al interferes with a proper absorption signal of Sr if both are present in a solution. The signal can be improved by the addition of La, which preferentially binds the interfering metal.

Ionization—If a large quantity of the test element is ionized, a very weak absorption is observed. This is due to the ionic absorption occurring at wavelengths different from that of the atomic one. The condition can be improved by adding a large excess of easily ionized elements; for example, in the measurement of Ca, a large amount of sodium ion usually is added.

Emission from the flame itself is minimized by using a chopper between the lamp and the flame. As the amplifier is designed to amplify only an ac signal (that of the chopping frequency), the intensity of light from the hollow cathode tube can be observed and recorded. A reduction of intensity due to the presence of the sample in the flame then will be detected. The magnitude of the decrease in intensity is a function of the quantity of the sample in the flame.

It is desirable to dissolve the sample in an organic solvent and for higher sensitivity the strongest absorption line must be chosen. In general, the resonance line resulting from the lowest excited state is usually the line exhibiting the strongest absorption. The instruction manual of each instrument suggests the choice of the line and the sampling technique.

New Atomization Techniques in AA Spectrometry

While flame atomic absorption still is used widely for the routine determination of more than 60 elements with new records of detection limits, there has been a considerable interest in the use of other atomization techniques. These include the mercury cold vapor atomic absorption, hydride generation techniques especially for As and Se, and an electrothermal atomization method. Recent research has concentrated on the latter technique as a new powerful tool for the study of trace amounts of lead in different matrices.

Fluorescence Spectrometry (Fluorometry)

When certain chemical substances are excited electronically by the absorption of UV or visible radiation, they emit light at a longer wavelength. This phenomenon is called *luminescence* and depending on the lifespan of the excited species, two different processes could be distinguished. The first is fluorescence, where the luminescence stops within 10^{-8} to 10^{-4} sec after the source of excitation is removed, and the second is phosphorescence, where the luminescence continues for a slightly longer period of time ($\sim 10^{-4}$ to 10 sec).

Theory

Upon absorption of visible or UV radiation by a molecule (usually $\pi \rightarrow \pi^*$ transition), the electron from S_0 (singlet ground state) is promoted to S_1 or S_2 (singlet excited states). The excited species, may return to the ground state by dissipation of energy through collision or by vibrational relaxation of the excited state. The vibrationally relaxed species can return to the ground state with the emission of radiation with a wavelength longer than that which originally was absorbed. This radiation is referred to as fluorescence. Figure 34-27 illustrates different electron spin states.

There also is a nonradiative process in which the excited state gives off energy and proceeds to a lower energy (triplet) state T, by a decay process. A return from T to S_0 gives off a long-lived radiation, which is called phosphorescence. The absorption and emission of radiation is specific for a particular molecule. Figure 34-28 is an energy level diagram that summarizes the electronic processes.

In order for a molecule to fluoresce, an absorbing molecular structure is required. Fluorescence may be expected to occur generally with molecules containing a highly conjugated system. At least one electron-donating group such as NH_2 or OH should be a part of the conjugated system. Electron
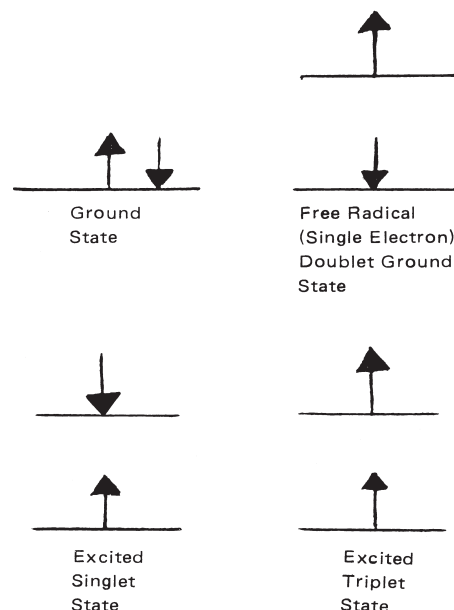
Figure 34-27. The different electron spin states.
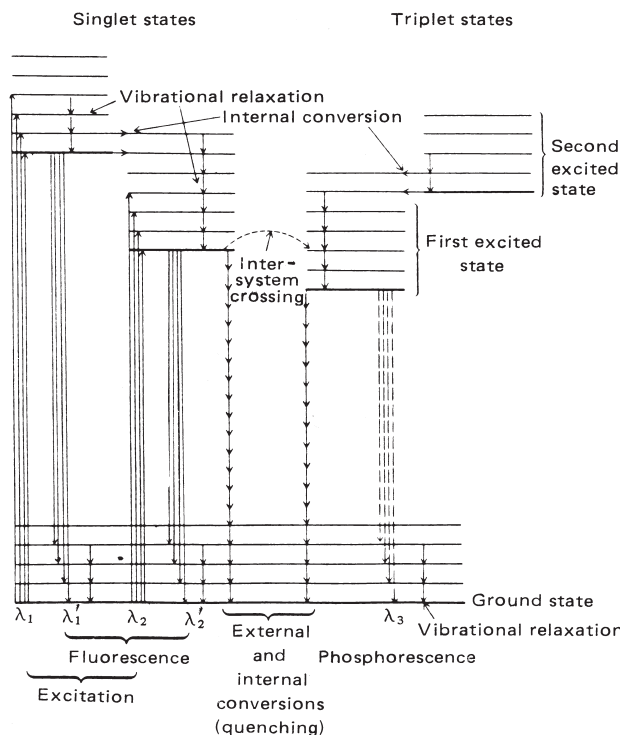
Figure 34-28. Energy level diagram for a photoluminescent system.[15]

withdrawing groups such as COOH or NO_2 diminish, and in some cases prevent, fluorescence. Fluorescence is enhanced as the rigidity of the molecule increases, a reduction in the internal vibration of the molecule.

In case of dilute atomic vapors, resonance fluorescence occurs (at the same wavelength as the excitation), however, in more complex organic compounds in addition to the resonance radiation, emission of radiation at longer wavelength occurs (Stokes' shift).

The position and intensity of the fluorescence bands are affected by pH. The quantum yield, φ, of fluorescence is lower than unity due to a "quenching" process; that is, not all of the excited molecules return to the ground state by emitting fluorescence radiation. Energy may be lost by bond dissociation and deactivation.

Fluorescence spectrometry offers detection limits lower than those of absorption spectrometry. A quantity of 1.1 μg/L can be measured and linearity can be maintained up to 10,000 μg/L. The method is applicable in the quantitative determination of fluorescing substances.

Fluorescence spectrometry has the greatest inherent sensitivity of all spectrometric techniques. Concentrations as low as $10^{-7}M$ can be measured accurately and precisely. It also has high selectivity, which makes it useful in the analysis of trace amounts of drugs and metabolites in biological fluids. Fluorescence, however, is used less widely than other absorption techniques due to the relatively limited number of organic compounds in which fluorescence can be induced.

Lately, because of computer enhanced techniques, new areas are being investigated such as *derivative fluorescence spectrometry*. In addition, fluorescence has proven to be of great value in HPLC where either natural or induced fluorescence can significantly lower the limit of detection. Furthermore, pre- and post-column derivatization of the sample to introduce fluorescence with such compounds as *o*-phthalaldehyde is becoming increasingly common.

Nonabsorptive Interaction of Matter with EM Radiation

The phenomenon of nonabsorptive interaction of matter with EM radiation is applied in such analytical procedures as light-scattering photometry, refractive index, and polarimetry. These interactions are not quantized, except for Raman spectroscopy, and therefore are considered to be nonspecific. However, each compound possesses its own characteristic interaction. The differentiation of stereoisomers with a polarimeter, the quantitative analysis of various substances with a refractometer, and the determination of the molecular weight of macromolecules by light scattering are examples of this type of instrumental analysis.

Light-Scattering Spectrometry

As in reflection and refraction, the scattering of radiation results when it passes through a transparent medium in which particles of a second phase are suspended. For EM radiation to be scattered by particles, two criteria should be met: the dimensions of particles should be equal to or smaller than the incident wavelengths, and the dispersing medium should have a refractive index different from that of the particles.

Particles of 0.1 to 1000 nm scatter EM in the UV and visible regions. If a beam of light is allowed to illuminate a colloidal suspension in a test tube, a pencil of light will be observed in the tube, due to the light-scattering phenomenon. This is known as the *Tyndall effect* and is an indication of the presence of suspended particles. The simplest kind of scattering is that observed by small, spherical and optically isotropic particles and is known as *Rayleigh scattering*.

Turbidimetry, nephelometry, and Raman spectrometry are analytical techniques based on the light scattering phenomenon. However, only Raman spectrometry will be discussed here.

Raman Spectrometry

In 1928 CV Raman, an Indian physicist, noted that under certain conditions, when an intense monochromatic light is scattered by molecules, the wavelength of a fraction of the scattered radiation is different from that of the incident beam. Figure 34-29 is a diagram of the various types of scattering of radiation. This shift (called the *Raman effect*) was found to be related to the chemical structure of the sample; therefore, it offered a new technique for structural elucidation and, in some cases, quantitative determination of several organic and inorganic compounds, in a way similar to IR spectroscopy.

Raman spectra, however, arise under certain conditions that are entirely different from IR. For example, molecules must undergo a change in their polarizability as they vibrate under quantum conditions, but are not required to have a dipole moment as in IR. Therefore, vibrations that are inactive in the infrared may be active in the Raman, such as homonuclear diatomic molecules. Also Raman spectra, unlike IR, can be used to study aqueous solutions.

The main limitation of Raman spectrometry, however, is that it is a weak effect, with low sensitivity and high vulnerability to much interference. Meticulous sample preparation is required, as any dust contamination would cause Tyndall scattering. Lately, laser sources, usually a helium-neon laser, have been employed to provide an intense, coherent, monochromatic beam, and this has improved the sensitivity significantly and raised new interest in the technique. Figure 34-30 is the Raman spectrum of carbon tetrachloride.
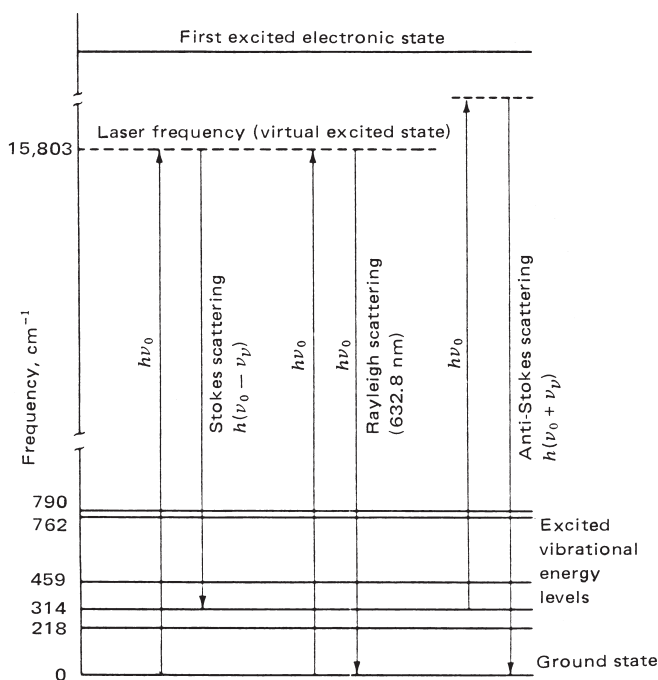
Figure 34-29. Energy interchange involved in Rayleigh and Raman scattering (CCl_4 molecule; source is a He-Ne Laser).[15]

Polarimetry

The fundamental principle of polarimetric analysis is based on the existence of optical activity in a substance, meaning the ability of a material to rotate plane-polarized light. Polarimetry is applicable to the determination of the molecular structure of substances that do not have a rotation–reflection symmetry axis. Determination of the sugar content of foodstuffs is an example of the quantitative application of polarimetry.

Modern polarimeters are capable of measuring optical rotation at more than just the traditional D-line of sodium. Some instruments measure optical rotation discretely at a number of different wavelengths. When the optical rotation is measured continuously as a function of wavelength, the technique known as *optical rotatory dispersion* (ORD) results. ORD has found some use in structural studies.

Solid-State Methods

Much must be accomplished in order to make an active pharmaceutical a viable product. The vast majority of pharmaceutical compounds are orally administered. The active molecule must be presented to the patient in a formulation in which the drug dissolves and is able to be absorbed through the small intestine. Furthermore, the drug must exist in a solid-state form that is sufficiently chemically and physically stable to produce the desired activity over a period of time, its shelf life, of typically two years or more. For this reason, solid-state chemistry and analytical methods play an important role in the development of the drug substance as both a bulk active pharmaceutical ingredient and as a formulated product.

Pharmaceuticals may exist in numerous solid forms having different physical properties. These forms include different salts of the pharmaceutical or neutral drug molecules. Each salt or neutral molecule may exist in number of physical states; it may exist in polymorphic forms, solvates and amorphous (non-crystalline) forms. As a consequence, a series of analytical techniques are employed that enable identification

of the existence of different forms, their characterization, and often quantification of mixtures of forms.[15–17] Ultimately one solid-state form is selected that provides the optimal physical properties. After the form is selected, analytical techniques are employed to ensure that the form that is in the final product is consistent with the form that was tested during clinical trials. Many of the spectroscopic techniques described previously, specifically solid-state NMR, FT-IR, and Raman spectroscopy are used for identification, by providing a fingerprint of the solid-state form. The differences in the spectra are due to perturbations of the molecular spectroscopy brought about by short-range molecular interactions such as differences in hydrogen bonding and conformation. Oftentimes the differences in the spectra are very subtle in spectroscopic methods. In contrast, x-ray diffraction patterns usually differ substantially from one crystalline form to another since diffraction probes the molecular organization within the three dimensional lattice, rather than perturbation of the molecular. Each technique has its own particular strength in solid-state characterization and as a consequence an integrated approach to characterization is recommended.[18]

Polarized Light Microscopy (PLM)

While it is not essential that molecules pack in an orderly crystalline environment, it is usually observed that drug substances do so. The crystalline environment provides a thermodynamically more favorable arrangement than does a disorderly "amorphous" form and has a higher, more efficient, density packing of molecules. As a direct result, crystalline arrangements of molecules typically give rise to more chemically stable drug substances, are less hygroscopic and give products that have better flow properties allowing for a more readily processed and formulated product. The simplest and most cost effective analytical technique used in pharmaceutical development is polarized light micrococpy (PLM).[19] For a fraction of the cost of other analytical techniques, PLM can be used to
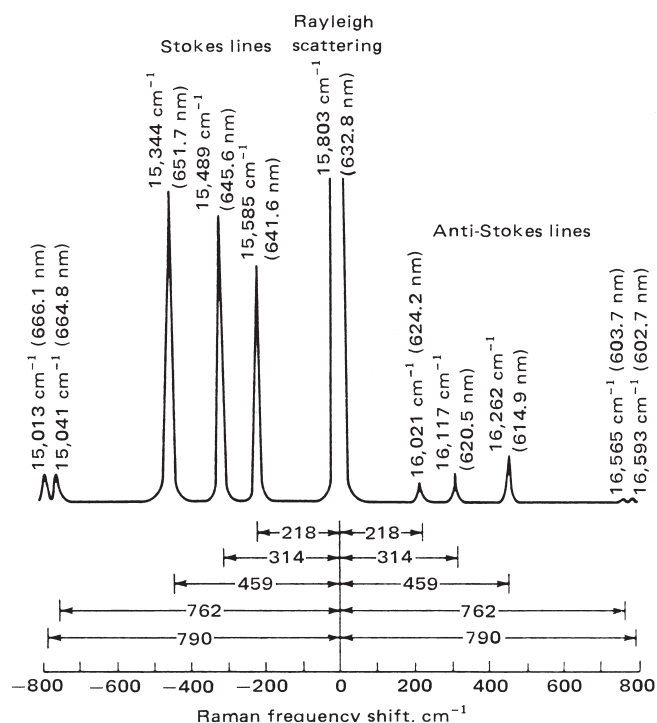
Figure 34-30. Raman spectrum of CCl_4 obtained with a He-Ne Laser.[15]

determine many physical properties of pharmaceutical compounds and plays a critical role in most laboratories due to its simplicity of use and the expediency with which information can be gained. Two important qualities that are instantly observed when one examines a solid are the presence of birefringence and the crystal habit or shape. The observation of birefringence under cross polarization indicates that the substance is crystalline, while crystal habit provides insight as to how well a material might process.

Theory

As was demonstrated earlier, molecules arranged in a crystal lattice are ordered and have a distinct orientation with respect to the facets of the crystal. Most organic molecular crystals crystallize in lower symmetry lattices; triclinic, monoclinic or orthorhombic lattices. Such lattices are termed anisotropic in that they possess more than one principal index of refraction and demonstrate the property known as double refraction. In double refraction, the light ray incident upon such a crystal is split into two components traveling at different velocities and in mutually perpendicular vibration directions. One can evaluate the crystalline quality of a sample by placing it between two polarizers and attempting to pass a beam of light through all three. Electromagnetic radiation can be represented as two vector components. As the ray passes through the first polarizer, only the fraction of the rays having vibrational direction parallel to the first polarizer, that is the vertical component, are transmitted to the crystalline sample. If the sample is an anisotropic crystalline solid, then this vertical component is again split into two components due to double refraction. As the rays strike the second, crossed polarizer, only the horizontal component is allowed to pass through to the eyepiece of the microscope, see Figure 34-31.

If the substance is isotropic, such as an amorphous solid, all of the light passing through the sample would remain unaltered (vertical) and would be removed by the second polarizer (PH).

The second rapidly assessed quality of a crystalline substances using microscopy is its external shape or "habit". Crystals that exhibit habits, where there is an elongated growth direction, can be problematic to develop due to poor flow properties and can significantly delay the speed with which a pharmaceutical can be developed.

The coupling of PLM with a hot stage can further extend the wealth of information that can be gained about a pharmaceutical solid.[20] It can be used to determine the melting characteristics and thermodynamic relationships between different crystal forms of a pharmaceutical. Polarized light microscopy can readily be automated using a precision X-Y sample stage and as result can easily be applied in high throughput crystal screening applications, as will be discussed later.

X-Ray Crystallography (Single Crystal)

X-ray crystallography provides the most detailed structural information about an organic molecule and its three dimensional packing in the crystalline lattice. Precise measurements are made that determine atomic connectivity, bond lengths, bond angles and torsion angles, thereby providing a complete de-
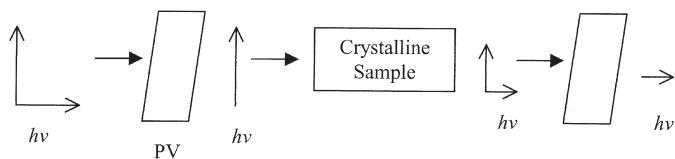
Figure 31. Illustration of birefringence with light passing through first a polarizer (PH) allowing verticle component to pass, then through an anisotropic crystalline solid followed by a polarizer (PH) that only allows the remaining horizontal component to pass to the eyepiece.

scription of the molecular conformation. X-ray crystallography is used to establish cis versus trans bond geometries and relative stereochemistry. It also enables one to establish absolute configurations of a molecule's chiral center(s) by relation to a known stereo-center of the molecule or by relation to a known stereo-center of a counter-ion forming a salt with the molecule. Even if a molecular structure contains only a single chiral center, its absolute stereochemistry can be established by the method of anomalous dispersion of a structure containing at least one heavy atom.[21] The greatest limitation of the technique is the requirement that material be isolated as a single crystal, having a smallest dimension on the order of 10 to 30 microns. Generally one skilled at crystallizations can meet this criterion in very short period of time.

Theory

The precise internal order of crystals can be demonstrated when a crystal is used as a three-dimensional diffraction grating for radiation that has a comparable wavelength to the interatomic distances within the crystal. In 1912, Max von Laue suggested the first diffraction experiments be conducted by Friedrich and Knipping to test the hypothesis that X-rays are wavelike with wavelengths on the order of 1 Å.[22] The success of the experiment led to the determination of the first crystal structure, NaCl, that same year by W.L. Bragg.[23] A single crystal is composed of millions of regularly organized molecules in unit cells that are stacked upon one another in a space filling arrangement. The smallest repeating unit, the unit cell, can be described by six lattice parameters; three axial dimensions a, b, and c and three angles α, β and γ. (see Table 9). The volume of the unit cell can be described by the general relationship

$$V = abc \, (1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma)^{1/2}$$

Or may have parameters constrained due to the symmetry to one of seven crystal classes as described in Table 34-9.

A distribution of the primitive and centering operations of lattice points among the seven crystal classes result in 14 different Bravais lattices, where equivalent lattice points are related by translational symmetry alone. When rotation, mirror reflection, and inversion symmetry operations of the 32 unique point groups are applied to the 14 different Bravais lattices, 230 unique space groups are possible for the organization of molecules to fill three-dimensional space. After the lattice parameters are determined by indexing, inspection of systematic absences are used to determine the space group. Systematic absences in the diffraction pattern indicate the presence of symmetry elements that cause selective and predictable destructive interference to occur. Tabulation of the systematic absence relationships, along with the unit cell dimensions during space group determination is done by systematically considering each class of Bragg reflections. Once the space group is deduced, attempts are made to solve the structure. For small molecule structures, composed primarily of organic molecules, direct methods have become the dominant method for structure determination.[24] The crystal structure provides the atomic locations of all atoms in the unit cell of the crystal, as is illustrated in Figure 34-32. Molecular structure information is provided to the food and drug administration as a part of the proof of structure section of the new drug application (NDA), whereas the information about crystallographic packing and the physical properties of the solid are often provided in the control and manufacturing section of the NDA.

Cryocrystallography

Data sets are commonly collected at reduced temperature. This, in effect improves the certainty with which atoms can be located. This is because there is a large contribution of vibrational mo-

Table 34-9. Seven Crystal Systems, Their Unit-Cell Geometries, and Bravais Lattices

CRYSTAL SYSTEM	NO. OF INDEPENDENT PARAMETERS	PARAMETERS	LAUE SYMMETRY	LATTICE
Triclinic	6	$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$	$\bar{1}$	P
Monoclinic	4	$a \neq b \neq c$; $\alpha = \gamma = 90°$; $\beta \neq 90°$	2/m	P, C
Orthorhombic	3	$a \neq b \neq c$; $\alpha = \beta = \gamma = 90°$	mmm	P,C,I,F
Tetragonal	2	$a = b \neq c$; $\alpha = \beta = \gamma = 90°$	$\underline{4}$/mmm	P,I
Rhombohedral	2	$a = b = c$; $\alpha = \beta = \gamma \neq 90°$	$\bar{3}$m	R
Hexagonal	2	$a = b \neq c$; $\alpha = \beta \neq \gamma$	6/mmm	P
Cubic	1	$a = b = c$; $\alpha = \beta = \gamma$	M$\bar{3}$m	P,I,F

tion to the anisotropic displacement parameters describing the electron density surrounding the atoms of a molecule. While cryo-crystallography increases the quality of the structural determination, there are other benefits. A very high percentage of organic molecular crystals contain the solvent of recrystallization, where the solvent molecule in the crystal lattice is water (hydrate) or another solvent molecule. Many solvents used for crystallization of organic compounds are volatile and are readily lost from the crystal lattice due to desolvation at room temperature. Desolvation results in fracturing of the single crystal into thousands of smaller "desolvated" crystallites. Temperature reduction stabilizes the solvated structure and enables the crystal structure to be solved.[25]

X-Ray Powder Diffraction (XRPD)

In addition to determining whether a compound is crystalline or amorphous, the first instrumental technique applied is typically x-ray powder diffraction (XRPD). XRPD enables the direct determination of the presence of different crystalline forms of a compound. XRPD is often referred to as the gold standard for determination of the existence of drug polymorphism. A compound is termed to be polymorphic when it crystallizes with different molecular arrangements but identical chemical constituents.[26] If one were to dissolve samples containing two different polymorphs, the samples are no longer distinguishable, that is they will have identical solution NMR, mass spectra, solution-state IR or solution-state Raman spectra.[15] Polymorphs have different physical properties due to their different packing arrangements. The different arrangements are responsible for differences in solubility, rates of dissolution, and melting points. Just as graphite and diamond have identical chemical compositions but different physical properties, so too do different crystal forms of a drug substance.

X-ray powder diffraction can be used to determine the physical state of a drug and used as a quality control assay to ensure that the drug is present in the proper crystallographic form so as to ensure reproducible performance of the active pharmaceutical ingredient in the drug product, hence control of the manufacturing process.

Theory

In the powder method, crystals to be examined are reduced to a fine powder and placed in a beam of monochromatic x-rays. Each tiny crystal is oriented at random with respect to the incident beam. The crystal diffracts x-rays similar to a diffraction grating. The three-dimensional crystal functions like a series of plane gratings stacked one above the other, giving rise to diffraction of electromagnetic ray whose wavelength approximate the atomic spacing in the crystal lattice.[27] The wavelength of the x-rays, λ, is related to the angle of incidence, θ, and the interatomic distance, d, by the Bragg equation;

$$n\lambda = 2d \sin(\theta)$$

where n is the order of the diffraction, 1, 2, 3, and so on.

The analogy of planes in a crystal owe their existence to the repetition of molecules packed in the crystal lattice, since diffraction occurs as the result of the interaction of the radiation with the electrons of the atoms. The planes are separated by an interplanar spacing of atoms, d, commonly termed the d-spacing. Constructive interference occurs when the path difference that the two rays travel are an integral number of wavelengths before they constructively recombine. When the "Bragg condition" is fulfilled, a peak is detected that is representative of the interplanar spacing of the symmetry equivalent sets of Miller planes. Figure 34-33 provides an illustration of cones of diffraction that emanate from a powdered sample when a beam of x-rays strikes the crystalline sample. The position that the
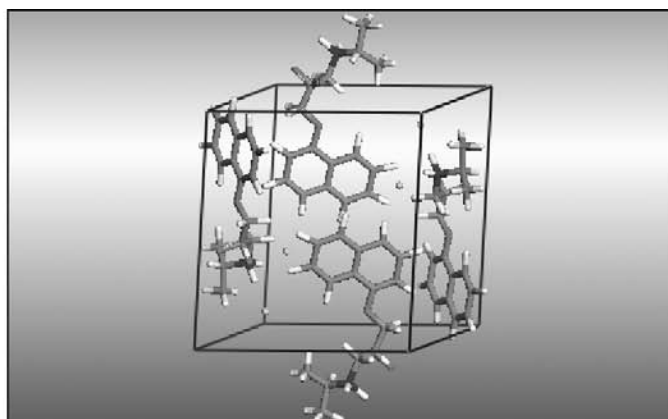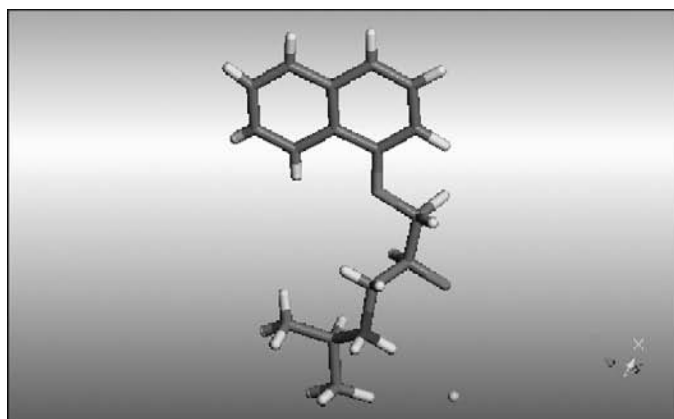
Figure 34-32. The 3-dimensional molecular structure of d,l-propranolol hydrochloride provides information about the molecular conformation and bonding whereas the its packing arrangement within the crystallographic unit cell is useful in understanding the physical properties of the crystalline form. See Color Plate 1.
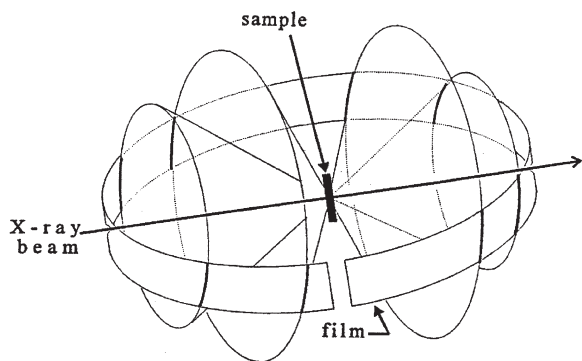
Figure 34-33. Cones of diffraction are produced by an x-ray beam striking a crystalline powder sample (reproduced with permission of Wiley-Interscience).[30]

Methodology

The x-ray diffractometer consists of an x-ray source that produces the x-rays, most commonly a sealed x-ray tube, a variety of optics that minimize divergence of the x-ray beam as it strikes the sample and travels toward the detector, see Figure 34-34.

In a typical powder diffraction experiment, the wavelength of radiation is constant. For the Bragg condition of diffraction to be satisfied, the angle of the diffracted beam that is detected must be the same as the angle of the incident x-ray beam. In the Bragg-Brentano geometry, this is accomplished by scanning the detector through an angle (2θ) while the sample is scanned through an angle (θ). An alternative approach requires movement of the heavier x-ray tube through an angle (θ) at the same rate as the detector is scanned through the angle (θ). Because of the historically more common (θ)-$2(\theta)$ geometry, powder patterns are usually depicted in a graph of intensity of diffraction (y-axis) versus the $2(\theta)$ angle of diffraction, see Figure 34-35.

diffracted radiation intersects with the detector, in this case a piece of photographic film, is characteristic of the material.

The peak positions of the unit cell reflect its size and angular relations of the crystal system. The intensity detected is a function of the atoms that make up the crystal and their scattering factors, a function of the electron density surrounding the atoms comprising the sample, as well as the location of the atoms within the unit cell. As a consequence, the diffraction pattern provides a unique characterization of a crystalline substance representing both its crystallographic packing and its unit cell contents. Each substance scatters the beam in a particular diffraction pattern, producing a unique fingerprint for each crystal form.

Instrumentation

Many different detectors exist that can be applied to diffraction, varying in speed of data acquisition, sensitivity, and resolution. Similarly there are many different sources for generation of x-radiation, having different characteristic wavelength and intensity. Most commonly a copper $K\alpha$ radiation source is used, having a characteristic wavelength of 1.54056 Å. Scintillation or solid-state point detectors are used to detect the diffracted radiation. There are numerous potential experimental errors that occur in and x-ray diffraction experiment. The major factor influencing observed intensity is the lack of random orientation of crystallites in the powder sample, whereas the major error in
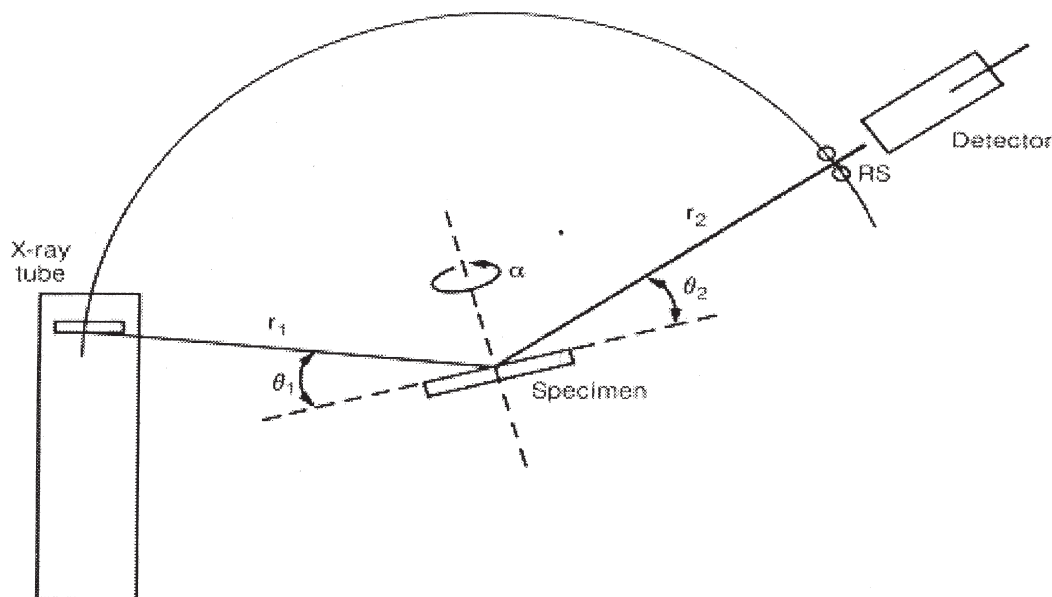
Type	Tube	Specimen	Receiving Slit	r_1	r_2
Bragg-Brentano θ:2θ	Fixed	Varies as θ *	Varies as 2θ	Fixed	$= r_1$
Bragg-Brentano θ:θ	Varies as θ	Fixed *	Varies as θ	Fixed	$= r_1$

Figure 34-34. Representation of the Bragg-Brentano geometry most commonly used the powder diffraction experiment uses a conventional x-ray tube source with a point detector (reproduced with permission Wiley-Interscience).[30]
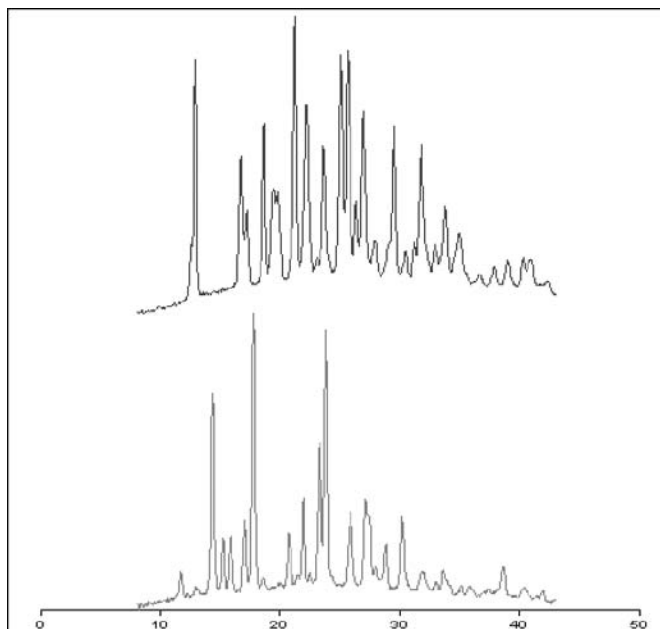
Figure 34-35. The x-ray powder diffraction patterns of two polymorphic forms of d,l-propranolol hydrochloride indicate differences in molecular arrangements within their different crystal lattices. See Color Plate 2.

angle of diffraction is due to sample displacement from the diffraction plane.[28–30] Using a properly aligned and calibrated instrument and appropriate sample preparation, these errors can be minimized.

Thermal Methods

Thermal analysis is a technique in which a physical property of a substance is monitored as a function of controlled temperature increase. Modern thermal analytical methods can measure weight loss on heating, melting points, heat and energy of tran-

sitions, and changes in form, in dimensions, or in the viscoelastic properties of the substance. They find wide applications in material characterization, purity determination of medicinal substances, study of relative heat stabilities and dynamic properties of new compounds, as well as in crystallography, chemical kinetics, and generation of phase diagrams.

Theory

Most thermodynamic events are accompanied by a loss of heat or require addition of heat from an external source to proceed. The event may be a phase transition, loss of a volatile component, or a chemical reaction. Each of these occurrences can be followed thermodynamically by noting either change of temperature of the sample under study or energy changes of the sample with respect to time. If the sample loses a volatile substance by evaporation, sublimation, or chemical conversion to a gas, it also is possible to follow the course of events by noting weight loss with respect to time, as the temperature of the sample is increased at a constant rate.

The general laws of thermodynamics, specifically those governing calorimetry, serve as the basis for understanding the theoretical concepts involved in the different thermal analytical methods of analysis. For equilibrium transitions, where $\Delta G = 0$ the heat of transition, ΔH_n is related to the entropy of transition ΔS_n by
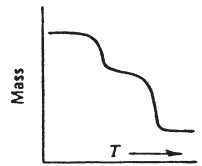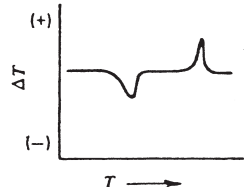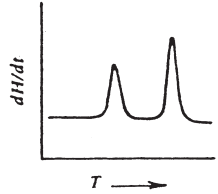
$$\Delta S_n = \frac{\Delta H_n}{T_n} \qquad (14)$$

Modern instruments for thermal methods of analysis are based on these parameters: mass, temperature, and heat flow. Table 10 illustrates the use of these functions and typical data outputs.

Thermogravimetry (Thermogravimetric Analysis, TGA)

Thermogravimetric analysis (TGA), perhaps the simplest form of thermal analysis, uses a *thermobalance* as the analytical instrument. The apparatus may be no more than a

Table 34-10. Typical Curves Produced in Thermal Gravimetric Analysis (TGA), Differential Thermal Analysis (DTA) and Differential Scanning Calorimetry (DSC)

TECHNIQUE	PARAMETER MEASURED	INTRUMENT EMPLOYED	TYPICAL CURVE
Thermogravimetry	Mass	Thermobalance	
Differential thermal analysis (DTA)	$T_s - T_t (\Delta T)$	DTA apparatus	
Differential scanning calorimetry (DSC)	Heat flow, dH/dt	Calometer	
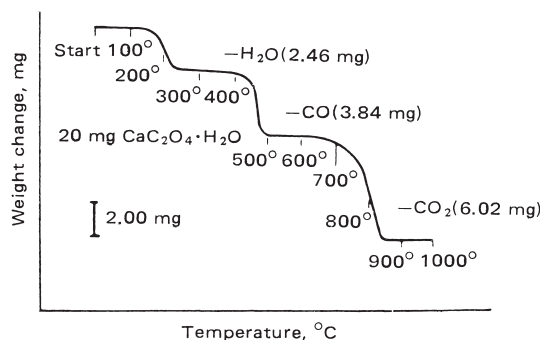
Figure 34-36. Thermogravimetric evaluation of calcium oxalate monohydrate, heating rate 6°/min.[19]

modified single pan analytical balance provided with a digital electronic output so that a plot of weight change (y-axis) can be made with respect to time or temperature. An infrared lamp may be the source of heat to irradiate the balance pan. Many modifications of such a device are used to determine the moisture content of tablet granulations, hydrated substances, and so on. Much more sophisticated instruments are also commercially available that include temperature programming and the use of a variety of beam, spring, cantilever, or torsion balances to determine changes in the weight of a sample. Because the atmosphere surrounding the heated sample may influence (retard or hasten) decomposition, provision often is made to control the atmosphere by addition of inert gases (nitrogen, helium) or reactive gases (oxygen, hydrogen, etc). The result of a thermogravimetric evaluation of calcium oxalate monohydrate may be seen in Figure 34-36.[31]

Recently several types of thermogravimetric devices have been coupled to a gas chromatograph, mass spectrometer, or FI-IR so that the effluent products of decomposition can thus be characterized.

A new development in TGA is the high-resolution technique.[32] The rate of heating of the sample is modified dynamically and continuously in response to changes in the sample decomposition so as to maximize weight change resolution. This technique allows use of very high heating rates while avoiding transition temperature overshoot, thereby optimizing time to complete a thermal analysis experiment.

Differential Thermal Analysis (DTA)

In differential thermal analysis (DTA), a sample and a thermally inert reference material are heated (or cooled) linearly with the aid of a programming device, and the temperature difference between the sample and the reference is measured as a function of the temperature applied. Because, during transition, the sample may either absorb or evolve heat, the difference in the temperature between the sample and the standard is equivalent to the temperature of transition and can indicate if the transition is endothermic or exothermic. Usually, ΔT is plotted against the temperature, T, or as a function of time (t). A block diagram of a typical differential thermal analyzer is depicted in Figure 34-37 and a schematic diagram of a modern DTA instrument is illustrated in Figure 34-38.[33]

DTA data are probably the most accurate of all thermal techniques, because the thermocouple is inserted into the sample; however, only the temperature of transition and not the amount of heat can be measured from a DTA curve, as the area under the peak is not proportional to the amount of energy transferred into or out of the sample.

Differential Scanning Calorimetry (DSC)

Another technique, very closely related to TGA, is DSC which differs only in that the sample and reference containers are not contiguous, but are heated separately by individual coils that are heated (or cooled) at the same rate. Platinum resistance thermometers monitor the temperature of the sample and reference holders and electronically maintain the temperature of the two holders constant.

If a thermodynamic event occurs that is either endothermic or exothermic, the power requirements for the coils maintaining a constant temperature will differ. This power difference (ΔP) is plotted as a function of the temperature recorded by the programming device.

Unlike DTA, in DSC the amount of heat put into the system is exactly equivalent to the amount of heat absorbed or liberated during a specific transition (transition energy).

AUTOMATION OF INSTRUMENTAL METHODS

Beginning in the early 1990s, a well-chronicled trend occurred in the pharmaceutical industry involving the implementation of high throughput screening (HTS) methods for lead generation. Today HTS assays involving in vitro assays, such as ligand-receptor binding and enzyme inhibition are routinely performed using highly automated schemes capable of processing over 10,000 samples per day. When combined with other related trends, such as combinatorial chemistry, parallel synthesis, and systems biology, the demand for sample preparation and analysis has transformed the modern pharmaceutical laboratory into a highly automated workplace. Unfortunately, a full review of laboratory automation is beyond the scope of this communication. Instead, this section will highlight select examples of technologies or practices that have increased throughput for the major instrumental techniques introduced in this chapter.

Although laboratory automation will not be the primary focus of this section, it is appropriate to mention the significance
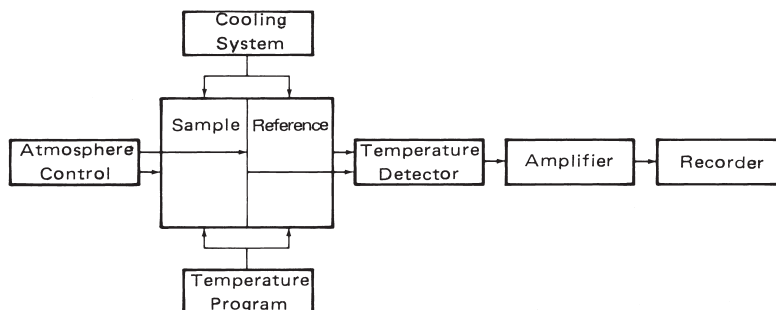
Figure 34-37. Block diagram of differential thermal analyzer.
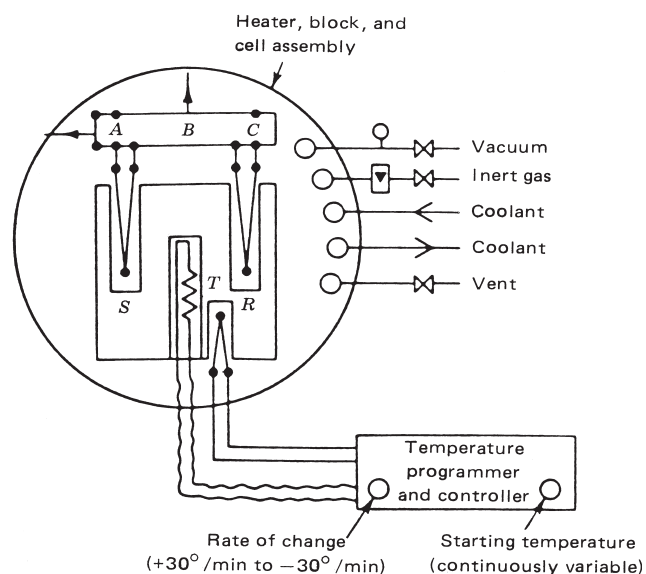
Figure 34-38. Schematic diagram of the DuPont differential thermal analysis apparatus.[17]

of the microtiter plate, which has become the default format for sample handling. Originally, introduced as a format for ELISA (enzyme-linked immunosorbent assay), the microtiter plate consists of 96 wells arranged in a standard format of 8 rows of 12 wells. In addition to offering a convenient approach for processing high sample numbers (compared to individual tubes), the microtiter plate provides a standardized platform for automated liquid handling and keeps reagent costs to a minimum. Today, the capacity for sample preparation and analysis is often described in terms of "plates" referring to the 96-well microtiter plate. Larger capacity formats based on multiples of 96 have also become popular (ie, 384, 1536).

In contrast to the HTS methods involved in lead generation, most of the instrumental techniques discussed in this chapter involve serial detection and are thus not capable of extremely high throughput (a notable exception is the spectrophotometric plate reader, which allows direct detection in a 96-well format). Due to the serial format inherent to many forms of instrumental analysis, extensive focus has be given to technologies which automate either sample *preparation* and sample *analysis*. Each of these topics is addressed separately below.

Sample Preparation

For many forms of instrumental analysis sample preparation simply involves dissolution in an appropriate solvent at the desired concentration. A common example is structural confirmation of individual compounds or compound libraries by MS and/or NMR. For such applications, the compounds undergoing analysis exist either as neat solids or concentrated solutions (eg, DMSO stock solutions). Given that compounds are typically processed using automated methods for liquid handling, most analytical instruments are equipped with autosamplers which accommodate 96-well plates.

Bioanalysis

In contrast to the relatively limited sample preparation associated with routine applications of organic structural analysis, bioanalytical applications tend to require more extensive sample preparation. By definition, bioanalysis refers to quantitative analysis of a drug and or its metabolites in a biological matrix and is most often linked to the study ADME properties.

Matrices studied can either be of *in vitro* (eg, liver preparation, biological media or buffer) or *in vivo* (eg, plasma, urine, tissue homogenate) origin. Because of its ability to perform rapid, selective quantitation in complex matrices, LC/MS/MS has become the default tool used to investigate ADME properties. ADME applications may also be qualitative, such as profiling and structural identification of drug metabolites in biological matrices. Although similar sample preparations are used for qualitative and quantitative applications, the present discussion will be limited to quantitative bioanalysis.

Despite the power of LC/MS technology, bioanalysis requires some form of sample preparation to achieve acceptable sensitivity, precision and robust instrumental performance. Several review articles have been published on this subject.[34, 35] The method selected for sample preparation depends both on the complexity of the sample matrix and the required sensitivity. Sample throughput and costs are also important considerations. For most drug discovery applications simple desalting or protein precipitation (PP) is employed to maximize throughput. PP, typically used for plasma and tissue bioanalysis, can result from the simple addition of an organic solvent such as acetonitrile or methanol. Other procedures involving acid or salt ($ZnCl_2$) may also be used.

Assays that require greater sensitivity generally involve some form of sample extraction to clean up and concentrate the sample. The two most widely adopted formats are solid-phase extraction (SPE) and liquid-liquid extraction (LLE). In SPE, the biological sample is loaded onto a solid sorbent having an affinity for the analyte. The SPE cartridge is then washed to remove matrix components followed by subsequent elution of the analyte. LLE involves placing the sample into a mixture of two immiscible solvents. Differential partitioning of the analyte and matrix components serves as the basis for sample cleanup. Both methods afford greater selectivity than PP and may also be used to concentrate the sample. All three aforementioned sample preparation methods routinely employ 96-well format and frequently incorporate automated procedures for liquid handling.[36–38]

Historically, LC/MS/MS methods were applied to development applications, such as GLP toxicology or clinical sample analysis. Over the past five years the demand for drug discovery bioanalysis has dramatically increased due to the use of LC/MS/MS for exposure screening as well as the widespread implementation of *in vitro* screens to assess ADME properties. Common in vitro ADME screens include hepatic metabolic stability,[39] intestinal permeation,[40] blood brain barrier penetration,[41] protein binding[42] assessment of drug-drug interaction potential[43] and the investigation of specific drug transporters.[44] Using LC/MS/MS technology, Cole and co-workers cite the ability to perform 2000 samples per instrument per day for a variety of in vitro ADME screens.[45] Further reading on this topic is available from a number of published articles.[46]

As with *in vitro* applications, increased demand for *in vivo* sample analysis has also occurred. *In vivo* demand stems from the important need to quantify drug exposure in live-phase studies ranging from early exposure screening in pharmacology animals to large clinical trials in man. In these studies, the principal sample matrix is plasma, since the plasma concentration of a drug is used to derive pharmacokinetic or toxicokinetic parameters. One of the issues faced in discovery bioanalysis is the need to develop rapid methods to simultaneously quantify several analogs in a structural series, as opposed to a single clinical candidate. This challenge is significant when strategies such as sample pooling[47] or cassette dosing[48] are used to maximize the utilization of expensive LC/MS/MS instrumentation. To accommodate the need for simultaneous quantification, many analysts employ gradient elution techniques, often in conjunction with one of several methods available which allow sample preparation and analysis to occur on-line. Note, traditional bioanalytical sample preparation techniques occur off-line (ie, independent of sample analysis). Reported approaches

to on-line sample cleanup include restricted access media,[49] turbulent flow liquid chromatography,[50] SPE[51] and immunoaffinity chromatography.[52] The common link to all of these approaches is the use of multiple columns, connected via multi-port HPLC valves, to derive an automated approach referred to as "column-switching." Under column-switching, a biological sample is loaded onto a dedicated extraction column that retains the target analyte while unwanted matrix components are washed to waste. After a defined period, the valve is switched to allow elution of the analyte onto a second column where chromatographic separation and detection are accomplished.

Perhaps the biggest difference regarding *in vivo* bioanalysis performed in drug discovery versus drug development is the sensitivity required. Discovery applications typically require quantification in the low ng/mL range, whereas development applications, such as the support of human clinical trials, often require quantification in the low pg/mL range. Nevertheless, a common thread is that LC/MS/MS is routinely used as the method of choice. Although many of the tools used for automation are common between discovery and development applications, the strategies used as well as the degree of regulatory oversight differ dramatically.

A final example of automated bioanalytical sample preparation is use of semi-automated systems for tissue preparation. Historically, tissue analysis is preceded by the homogenization of individual samples, a process that is both time and labor-intensive. Recently, a commercial instrument was introduced, which allows homogenization to be carried out in parallel by a series eight homogenization probes. The fully-automated system contains a platform, which accommodates 48 samples and has three wash stations to reduce inter-sample carry over. The demonstration of this system for bioanalysis of brain tissue samples has been recently published.[53]

Automated X-ray Diffraction

One of the most dramatic examples of automated spectroscopic sample preparation is in the field of x-ray diffraction (XRD). For years, XRD has been regarded as the definitive method for the determination of protein structure; however, the throughput of the technique has been limited by the ability to grow acceptable protein crystals. In addition to taking several days, protein crystallization still requires a significant degree of 'trail and error' as several permutations in conditions must be tested.

A major breakthrough occurred when it was discovered that viable protein crystallization can occur using small protein sample volumes ($< 1 \mu L$).[54] In addition to requiring less protein, such nano preparation methods reduce the time needed for crystallization by as much as 10-fold. Using automation, it is possible to test hundreds of crystallization methods per day using automated schemes. While several permutations of this technology exist,[55,56] most approaches employ a variation of a technique referred to as "hanging drop vapor diffusion".[56] This technology has been employed to direct the synthesis of compound libraries in the generation of more specific leads. Co-crystallization of lead compounds with the target protein is a secondary technique applied later during lead optimization.

Sample Analysis

Most forms of instrumental analysis provide the capability for unattended sample introduction, a process typically referred to as automated sample introduction or "autosampling." Moreover, sample introduction can either occur by direct means or through an on-line interface to one of several forms of chromatography. To date, chromatographic interfaces exist for following spectroscopic techniques: MS, UV, fluorescence, IR, NMR, and electrochemical. Obviously, a review of "hyphenated" chromatographic techniques is beyond the scope of this section. Instead, attention will be given to current trends in sample introduction associated with the major forms of spectroscopy.

NMR

Conventional NMR sample introduction occurs by inserting a narrow glass tube containing a liquid sample into the center of a superconducting magnet. For years, unattended sample analysis has occurred using robotic arms standard on all instruments which insert and retrieve the glass sample tubes from the magnet. As discussed previously, NMR flow probes have been introduced to permit coupling with HPLC. This same device is also used for flow injection analysis (FIA) providing another convenient means for automated, higher throughput sample analysis.[57]

NMR holds the distinction of being the first spectroscopic method to perform routine analysis in a mode known as "open access" (OA). Under OA, multiple users are able to perform standard analysis in a "walk-up" mode. In this environment, individual users place their samples in an autosampler, where they are queued for analysis in the order received. Many variations of OA exist, generally differing by the degree of instrumental control afforded to the walk up analyst. As one would expect, OA provides the most immediate turn around for sample analysis and works best when generalized analysis conditions can be employed. It is also important that the instrumentation be sufficiently robust to permit access by multiple users, often with limited training.

Mass Spectrometry

Routine application of MS in an OA environment did not occur until the introduction of the API-interface. The first reports of OA by MS occurred in 1994 and used APCI as the ionization mode.[58] ESI methods for OA have also been reported.[59] Both techniques are widely used in the pharmaceutical industry to support chemical synthesis. Although the most frequent format for OA sample introduction is FIA, it is important to note that OA by LC/MS is also a routine tool.[60]

Several approaches have been used to contend with the limitation imposed by the serial nature of MS detection. Using a novel redesign of a commercial autosampler, Morand and co-workers were able to perform FIA analysis in support of chemical synthesis at a rate of 3 seconds/injection (5 minutes/plate).[61] Several formats for ultra fast gradient elution have been employed with LC/MS[62] including the use of monolithic column formats.[63] For bioanalytical applications, full gradient elution routinely occurs with run times under 2 minutes. Staggered injections have also been used to increase LC/MS throughput.[64] This general methodology is applicable whenever analyte detection represents a finite part of the overall injection duty cycle.

Another approach taken is to combine multiple ESI sprayers with a single instrument. A commercial version of this methodology, referred to at MUX, employs a rotating cylinder containing an aperture to sequentially sample up to eight sprayers housed within a single API-MS interface.[65] Using the eight-sprayer format, each sprayer is sampled every 1.2 s allowing for multiple points to be acquired over a given chromatographic peak. MUX interfaces are used in conjunction with multi-injector autosamplers, where typically a gradient formed by a single HPLC system at elevated flow is prior to the autosampler to provide flow for multiple LC columns. Successful applications of this approach have been reported for both chemical synthesis support[66] and bioanalysis.[67]

An emerging trend in instrumental analysis is miniaturization. Often referred to as "lab-on-a chip," microfluidic applications, allowing sample analysis and detection to occur on a microchip, offer several advantages including reduced reagent costs and lower waste stream production. Current formats and applications for lab-on-a chip exist and have been
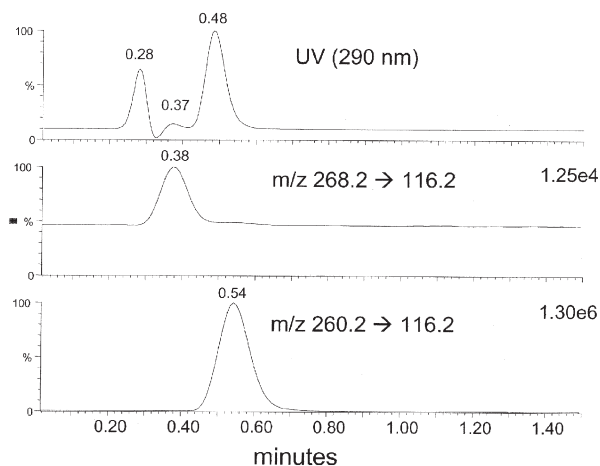
Figure 34-39. LC-UV-MS/MS anlysis of β-blocker drugs propanolol and metoprolol

reviewed.[68] The most common detection formats for microfluidics are spectrophotometric and include UV[69] and fluorescence.[70] Electrochemical detection has also been employed.[71] Recently, a silicon chip incorporating 100 individual ESI-MS nozzles was commercially introduced.[72] To date, this technology has been applied to a variety of applications including proteomics,[73] protein-ligand interactions[74] and bioanalysis.[75]

Application

Earlier in this chapter, the drug propranolol was introduced to illustrate the power of NMR for structural analysis. In the section that follows, further analysis of propranolol was conducted to provide additional examples of how common instrumental techniques are used for pharmaceutical analysis.

LC/M/MS Analysis of Propranolol

Due to its versatility, LC/MS has become one of the most widely used instrumental techniques for pharmaceutical analysis. LC/MS has the unique ability to extract structural information from complex mixtures with high sensitivity. Because of these attributes, LC/MS has been applied to several areas related to chemical synthesis. In early drug discovery, LC/MS is routinely

used to confirm the structure of newly synthesized molecules and to obtain estimates of purity. Figure 34-39 displays data obtained from the LC-UV-MS/MS analysis of a binary mixture of two β-blockers, propranolol and metoprolol. In this example, LC/MS was conducted by reversed phase HPLC with on-line UV detection followed by positive ion electrospray ionization (ESI). Gradient elution was conducted using a mobile phase system consisting of methanol, water and formic acid at a flow rate of 0.25 ml/min. The column used was an Aquasil C18 (2.1 mm × 2 mm). In this example, the ratio of propranolol to metoprolol was 200:1.

The upper profile in Figure 34-39, which corresponds to the UV absorbance at 290 nm, contains three peaks at 0.28, 0.37 and 0.54 minutes. The two most highly retained peaks correspond to the metoprolol and propranolol, respectively, whereas the peak at 0.28 minutes represents the column void (ie, unretained material). The relative insensitivity of UV is apparent from the metoprolol peak, which was barely detected. The signal shown corresponds to approximately 0.625 ng injected on-column for metoprolol versus 125 ng for propranolol. The corresponding MS signals were acquired by tandem MS using SRM detection (see mass spectrometry section). The bottom profile, corresponding to metoprolol, indicates the superior sensitivity of MS detection. The results in Figure 34-39 reveal a slight shift in retention time between UV and MS. This delay is related to the transit time between detectors. The incorporation of spectroscopic detection on-line with MS is applied to several forms of pharmaceutical analysis including chemical synthesis, drug metabolism, natural product analysis, stability testing and impurity profiling.

As described in the section on mass spectrometry, SRM detection involves monitoring a compound-specific fragmentation transition, typically using a triple quadrupole mass spectrometer, for the purpose of enhancing detection. Because of the extraordinary selectivity conferred by this approach, SRM vastly improves the signal to noise for analytes detected in a complex matrix and is therefore routinely used for quantification of drugs and their metabolites in biomatrices (eg, plasma, tissue homogenate, urine). Figures 34-40 and 34-41 display the product ion mass spectra for propranolol and metoprolol, respectively. The structure of each molecule appears as an inset in the corresponding spectrum. In each case, the protonated molecule was selected by Q1 and induced to fragment by collisions with an argon target gas in the Q2. The data shown were obtained by scanning the third quadrupole (Q3) to transmit the product ions formed in Q2. As indicated in Figures 34-40 and 34-41, both drugs gave rise to a prominent fragment ion at m/z 116, formed by the loss of the corresponding phenol as a neutral molecule. This fragmentation transition was optimized and used to acquire the SRM data
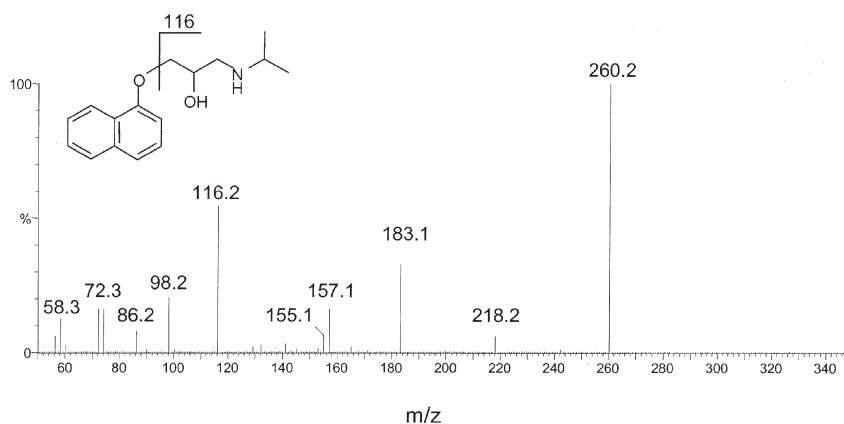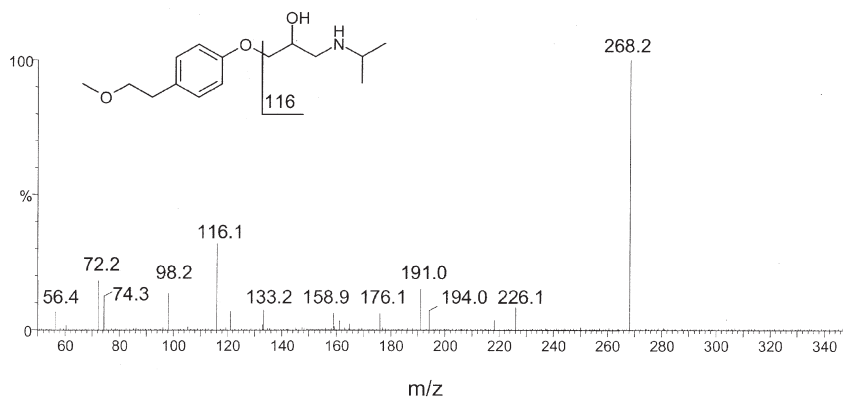

Figure 40. Product ion mass spectrum of propanolol

Figure 34-41. Product ion mass spectrum of metoprolol

presented in Figure 34-39. The corresponding SRM transitions used were as follows: propranolol (m/z 260.2 to 116.2) and metoprolol (m/z 268.2 to 116.2).

An illustration of how SRM detection is used for quantitative analysis is indicated by the calibration curve shown in Figure 34-42. To produce this curve, a series of neat propranolol standards were prepared and analyzed covering a range from 10 to 5000 ng/ml. All samples received a constant amount of metoprolol (25 ng/ml) acting as an internal standard. The calibration curve in Figure 34-42 plots the peak area ratio of propranolol to metoprolol versus propranolol concentration. Least squares linear regression analysis with 1/X weighting was used to fit a straight line through the data. The observed coefficient of determination (r^2) was 0.9978.

High Throughput Screening for Polymorphism of d, l-Propranolol Hydrochloride

At the very early stages of lead optimization, a vast number of molecules exist that demonstrate activity against target receptor sites. The number of molecules being evaluated and rate of attrition is high. In an effort to increase the rate that new compounds can be progressed from discovery to launch, a need has emerged to rapidly screen molecules for solid-state forms that posses acceptable biopharmaceutical and solid-state properties. The primary objective is to conduct meaningful toxicological studies on a solid form that can readily be developed and brought to market.
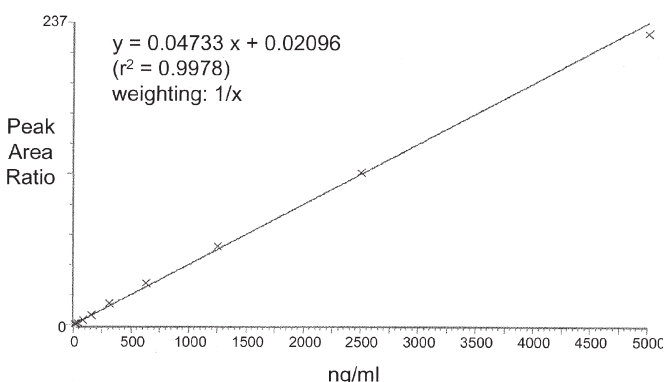
Figure 34-42. Standard curve for propanolol acquired by LC/MS/MS.

There are numerous counter-ions available for salt formation, particularly for actives containing basic moieties.[76] There are also numerous experimental variables that can result in successful crystallization. Three of the most common methods for crystallizing organic molecules are evaporation, precipitation by anti-solvent addition, and temperature reduction. The need for high throughput crystallization screening stems from the need to rapidly survey a large array of chemistries and crystallization variables in parallel. Critical to this end is the coupling of rapid experimental design with liquid and sample handling, followed by delivery of the isolated crystals in an arrayed format such that each individual sample can be characterized for desirable solid-state properties.

Central to the overall process is a database that provides the ability to track the chemistry and conditions and ability to rapidly analyze the data so that correlations can be made between the variables and crystallization outcomes.

Figure 34-43 provides the general strategy for crystallization screening of an active pharmaceutical. The same basic workflow design is used for salt formation or polymorph screening.

Design

The chemistry is designed such that a small amount of API is dispensed into individual wells and reacted with a stoichiometric amount of an acceptable counter ion. In the design, typically a 96 well plate is designed with the different counterions dispensed along a given column and different solvent or solvent combinations are dispensed along the rows, typically screening 12 counterions versus 8 different solvent combinations. In the case of a polymorphism, no counterion is dispensed. Each of the 96 wells has the same chemistry, but different solvent compositions. The design for crystallization screening of polymorphism screen is depicted in Figure 34-44.

Solvents or solvent mixtures for crystallization are dispensed, using an automated liquid handler. It is essential that every crystallization attempt be a unique experiment and that it be unbiased by the presence of preexisting nuclei. In order to accomplish this, the solution is hot filtered and the filtrate is transferred to each of three crystallization plates using a heated syringe and heated filtration assembly. One 96-well plate contains a predispensed, solvent-miscible antisolvent for precipitation of the solid. The second plate is maintained at elevated temperature and is cooled gradually to sub-ambient temperatures. The third dispense is to a 96-well plate where the sample is allowed to evaporate. Evaporation rate can be controlled at any set temperature.

An aliquot of sample is transferred initially to another 96-well plate to determine the solubility at elevated temperature
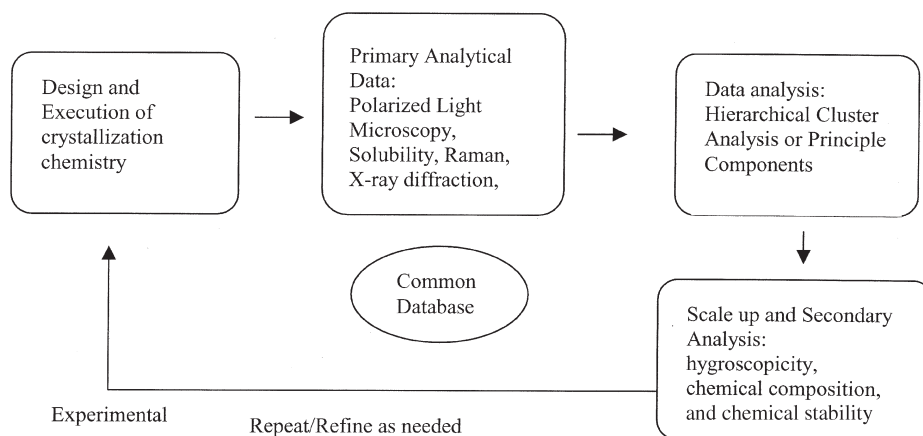
Figure 34-43. The design of the general workflow used for high throughput crystallization and solid-state characterization.

and another sample is transferred from the plate containing the sample that was cooled gradually, typically 55 and 10°C respectively. This sample is analyzed by HPLC or UV/VIS spectroscopy to provide an approximate measure of solubility. The solubility data can be useful in designing the next series of screening experiments.

Analysis of Experiments

One of the key elements working with small sample sizes is elimination of the transfer of the sample from one technique to another for analysis. Such transfers result in loss of sample. Central to the overall process is the ability for instrumentation to receive and examine samples in an arrayed, typically 96 well, format. The general procedure involves the use of a glass substrate on which the crystallization occurs. Birefringence assessment using polarized light microscopy and Raman microscopy and x-ray powder diffraction analysis is conducted directly on the glass substrate.

In the workflow for HTS crystal form we use polarized light microscopy to first assess whether or not a sample is crystalline and then use both x-ray powder diffraction and Raman microscopy to determine the number of forms present. The complementary methods for form assessment are used, since both techniques have advantages and disadvantages depending upon the particular sample. Figure 34-45 provides an illustration of the birefringence images and powder diffraction patterns from one of the three crystallization plates obtained in the HTS of d,l-propranolol hydrochloride, as can readily be seen there are at least two crystalline polymorphic forms. The plate layout, with each form is represented as a thumbnail whose color indicates its form, enables easy correlation of the crystallization outcome with the chemistry of the specific well.

After the different crystalline "hits" are identified, the chemistry information is used to try to validate the hits by scaling up the individual forms. Using the larger quantity of sample that is generated, other characterization tests are used to determine which form is best suited for further development. For crystalline salt forms, the stoichiometry of the salt is determined
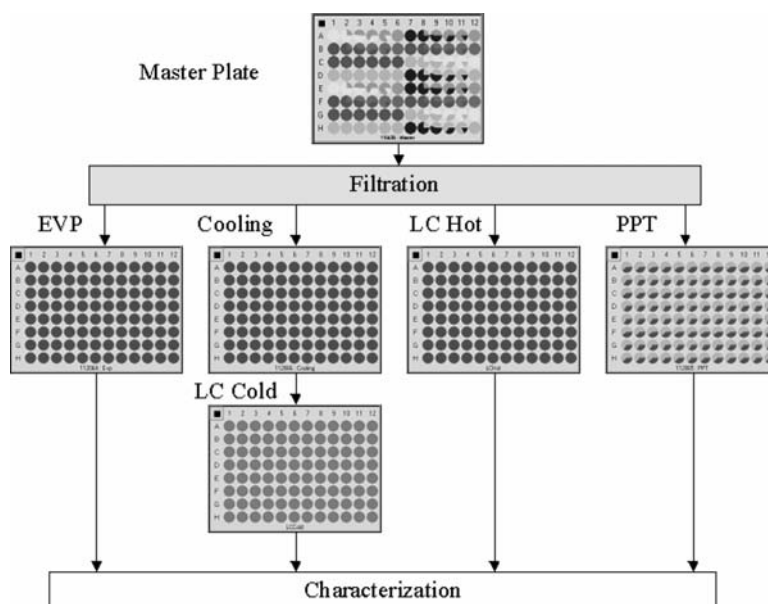
Figure 34-44. Design for the crystallization process for polymorphic form screening demonstrates hot filtration of the crystallization solution and its transfer to three crystallization plates and two plates for solubility determination. See Color Plate 3.
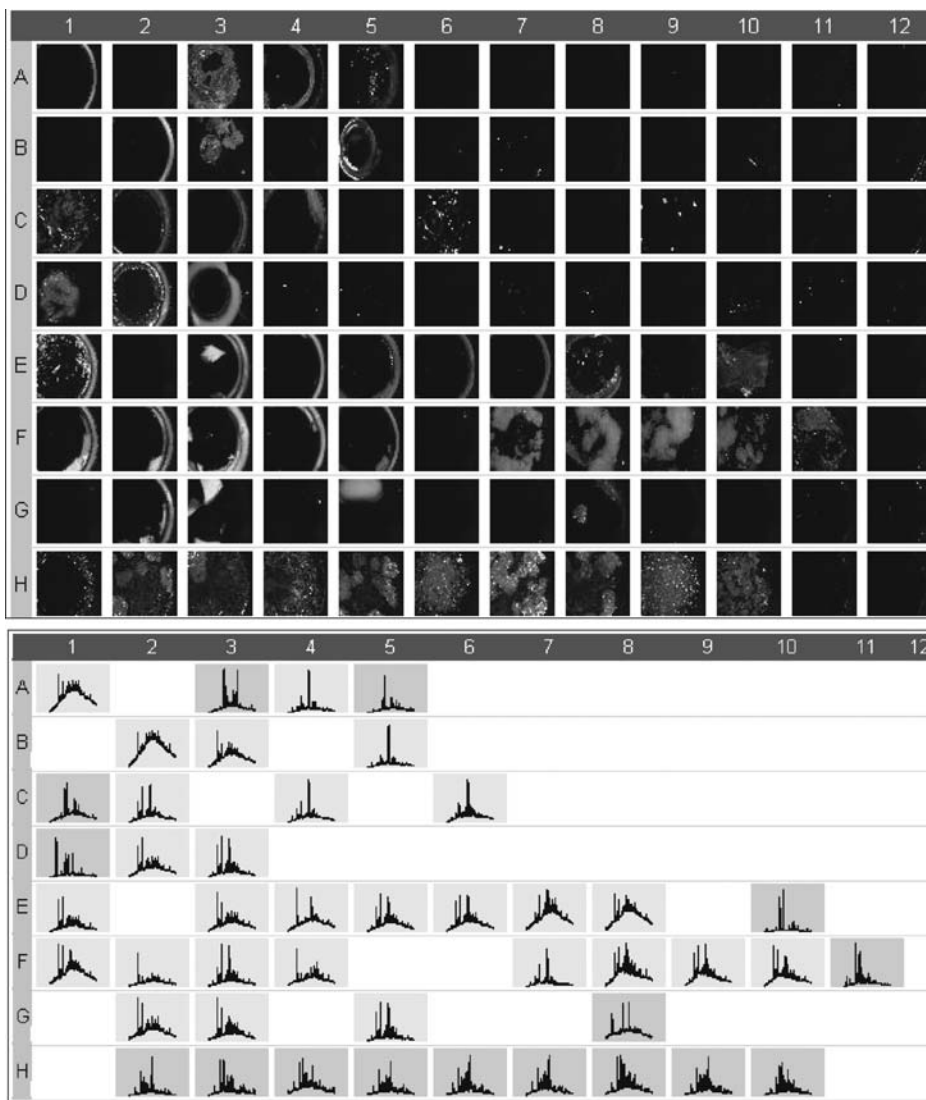
Figure 34-45. Birefringence images and powder diffraction patterns collected from the evaporative crystallization plate in the HTS of d,l-propranolol hydrochloride indicates two polymorphic crystal forms and their location within the 96-well plate, thus enabling correlation of crystallization chemistry with the crystal form obtained. See Color Plate 4.
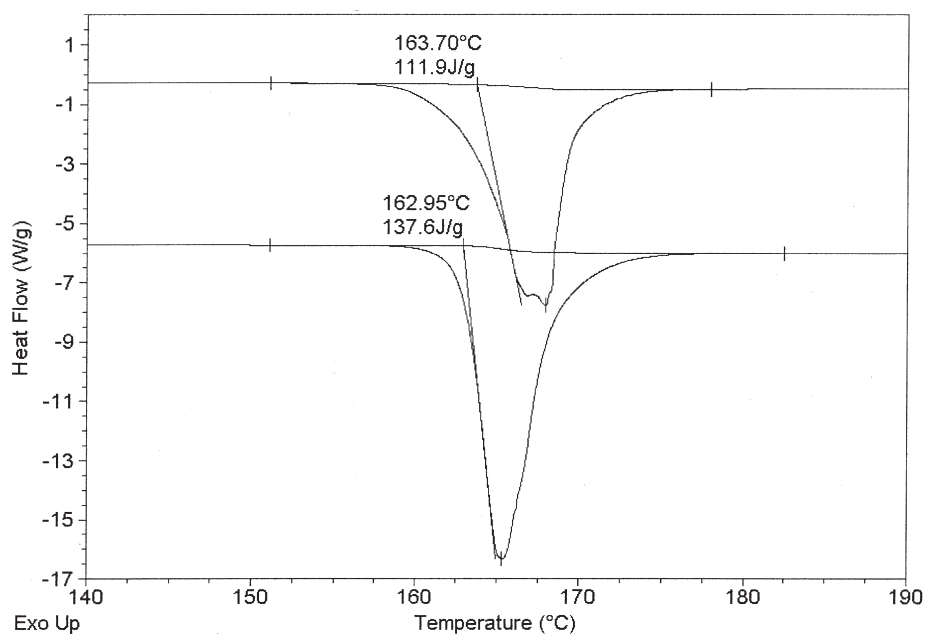
Figure 34-46. The DSC thermal curves of from two polymorphic forms of d,l-propranolol hydrochloride noting the onset temperature of melting and the enthalpy of fusion.

Propranolol
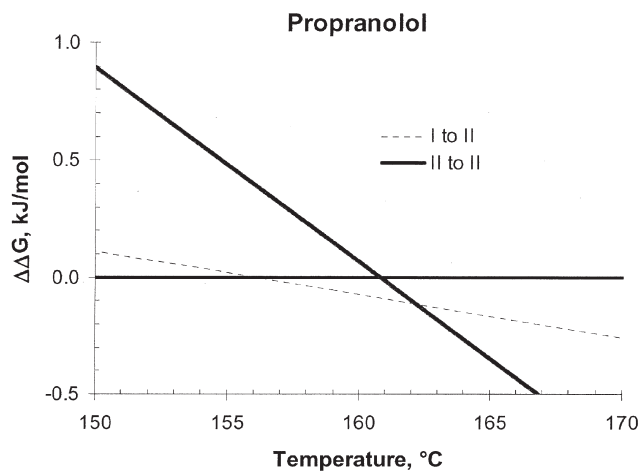
Figure 34-47. A temperature stability diagram of the two forms demonstrates that below 159°C the polymorphic form with the lower melting point is more stable whereas above the transition point, the higher melting polymorph is more stable.

using methods such as ^{13}C NMR or ion chromatography. The solid-state chemical stability of the forms is determined. The hygroscopic characteristics of the forms are assessed to ensure that the compound can be processed under normal ambient conditions. The thermodynamic relationship of polymorphic forms is determined so as to ensure that the most stable form under ambient temperature is produced. There is always a risk that a metastable form may eventually be unattainable during crystallization due to generation of a more stable. There have been numerous reports in which a metastable polymorphic form that was initially isolated could no longer be produced due to the appearance of a thermodynamically more stable form. On one occasion, the appearance of a more stable crystalline form of ritanovir, post product launch, resulted in a major product recall and an additional year of development due to the discovery of and the innovator's inability to isolate the crystalline form that was tested in clinical trials. One analytical method that is often used to determine crystal form stability is DSC. If the enthalpy of fusion of the higher melting form is greater than the lower melting form, then it is more stable at all temperatures between 0 K and its melting point and the polymorphs are related monotropically. If the difference in free energy versus temperature curves cross, then they are enantiotropically related and there is a temperature at which the lower melting polymorph is more stable than the higher melting polymorph, see Figure 34-46. In this trace it is evident that the higher melting polymorphic form has the lower enthalpy of fusion. The information indicates that the polymorphic forms of d,l-propranolol hydrochloride are enantiotropically related, with one melting at 162.95°C and the other at 163.7°C. That is the free energy of forms cross due primarily to differential changes in entropy with respect to change in temperature. A diagram of their stability versus temperature indicate that the transition temperature is 159°C, as is shown in Figure 34-47.

Acknowledgments

REFERENCES

1. Freeman R, Morris GA. *Bull. Magn. Reson.* 1979; 1: 5.
2. Lauterbur PC. *Nature (London)* 1973; 242: 190.
3. Hurd, R.E. *J. Magn. Reson.* 1990; 87: 422.
4. Shaefer J, Stejskal EO. *J. Am. Chem. Soc.* 1976; 98: 1031.
5. Crouch RC, Martin GE. *J. Nat. Prod.* 1992; 55: 1343.
6. Keifer PA, Baltusis RDM, Tymiak AA, Shoolery JN. *J. Magn. Reson.* 1996; A 119: 65.
7. Albert K. *J. Chromatogr.* 1995; 703: 123.
8. Shockcor JP, Silver IS, Wurm RM, Sanderson PN, Farrant RD, Sweatman BC, Lindon JC. *Xenobiotica.* 1996; 26: 41.
9. Ehlhardt WJ, Woodland JM, Baughman TM, Vandenbranden M, Wrighton SA, Kroin JS, Norman BH, Maple SR. *Drug Metab Dispos.* 1998; 26: 42.
10. Olson DL, Norcross JA, O'Neil-Johnson M, Molitor PF, Detlefsen DJ, Wilson AG, Peck TL. *Anal Chem.* 2004; 76: 2966.
11. Logan TA, Murali N, Wang G, Jolivet C. *Magn. Reson. Chem.* 1999; 37, 512.
12. HP8452A *Diode-Array Spectrophotometer Handbook.* Palo Alto, CA: Hewlett Packard, 1990.
13. Skoog DA, Holler FJ, Nieman TA. *Principles of Instrumental Analysis, 5th* Ed. New York: WB Saunders, 1998.
14. Willard HH, Merritt LL, Dean JA, Settle PA. *Instrumental Methods of Analysis, 6th Ed.* New York: Van Nostrand, 1981.
15. Haleblian JK, McCrone W. *Journal of Pharmaceutical Sciences* 1969; 58: 911.
16. Byrn, SR, Pfeiffer RR, Stowell JG. *Solid-State Chemistry of Drugs,* 2nd ed. SSCI, Inc, West Lafayette IN, 1999.
17. Stephenson GA, Forbes RA, Reutzel-Edens SM. *Advanced Drug Delivery Reviews* 2001; 48: 67.
18. Yu L, Reutzel SM, Stephenson GA. *Pharmaceutical Science and Technology Today* 1998; 1: 118.
19. Bloss, FD. *An Introduction to the Methods of Optical Crystallography,* Saunders College Publishing, Marietta, OH, USA, 1961.
20. McCrone WC. *Fusion Methods in Chemical Microscopy,* Interscience Publishers, New York, USA, 1957.
21. Giacovazzo C, Monaco HL, Viterbo D, Scordari F, Gill G, Zanotti G, Catti M. *Fundamentals of Crystallography,* ED C. Giacovazzo. Oxford University Press. 165, 1992.
22. Friedrich W, Knipping P, Laue M. *Sitzb. Kais. Akad. Wiss.,* Munchen, 1912; 303: 22.
23. Bragg, WL. *Proc. Roy. Soc. London* 1913; (A) 89: 248.
24. Glusker JP, Lewis M, Rossi M. *Journal of Chemical Education* 1995; 72: A73.
25. Hope H. *Acta Crystallographica, Section B: Structural Science* 1998; B44: 22.
26. McCrone WC. *Physics and Chemistry of the Organic Solid State* 1965; 2: 725.
27. Bragg WL. *Proc. Roy. Soc. London* 1913; (A) 89: 248.
28. Klug HP, Alexander LE. *X-Ray Diffraction Procedures*; CAPLUS 716, 1954.
29. Cullity BD. *Elements of X-Ray Diffraction.* 2nd Ed., CAPLUS 555, 1978.
30. Jenkins R, Snyder R. eds *Introduction to X-Ray Powder Diffractometry.* CAPLUS 544, 1996.
31. Strobel HA, Heineman WR. *Chemical Instrumentation: A Systematic Approach,* 3rd ed. New York: Wiley, 1989.
32. High Resolution Option Manual. New Castle, DE: Y/A Instruments, 1991.
33. Willard HH, Merritt LL, Dean JA, Settle PA. *Instrumental Methods of Analysis,* 6th ed. New York: Van Nostrand, 1981.
34. Ackermann BL, Berna MJ, Murphy AT. *Curr. Top. Med. Chem.* 2002; 2: 56.
35. Jemal M. *Biomed. Chromatogr.* 2000; 14: 422.
36. Allanson, JP, Biddlecombe RA, Jones AE, Pleasance S. *Rapid Commun. Mass Spectrom.* 1996; 10: 811.
37. Stenborner S, Henion J. *Anal. Chem.* 1999; 71: 2340.
38. Watt AP, Morrison D, Locker KL, Evans DG. *Anal. Chem.* 2000; 72: 979.
39. Di L, Kerns EH, Hong Y, Kleintop TA, McConnell OJ, Huryn DM. *J. Biomol. Screen* 2003; 8: 453.
40. Caldwell GW, Easlick SM, Gunnet J, Masucci JA, Demarest K. *J. Mass Spectrom.* 1998; 33: 607.
41. Chu I, Liu F, Soares A, Kumari P, Nomeir AA. *Rapid Commun. Mass Spectrom.* 2002; 16: 1501.
42. Fung EN, Chen YH, Lau YY. *J. Chromatogr. B. Analyt Technol Biomed Life Sci.* 2003; 795: 187.
43. Dierks EA, Stams KR, Lim HK, Cornelius G, Zhang H, Ball SE. *Drug Metab Dispos.* 2001; 29: 23.

44. Chen C, Hanson E, Watson JW, Lee JS. *Drug Metab Dispos.* 2003; 31: 312.
45. Janiszewski JS, Rogers KJ, Whalen KM, Cole MJ, Liston TE, Duchoslav E, Fouda HG. *Anal. Chem.* 2001; 73: 1495.
46. Kerns EH. *J. Pharm. Sci.* 2001; 1838.
47. Hop CECA, Wang Z, Chen Q, Kwei G. *J. Pharm. Sci.* 1998; 901.
48. Berman J, Halm K, Adkison K, Shaffer J. *J. Med. Chem.* 1997; 40: 3.
49. Needham SR, Cole MJ, Fouda HG. *J. Chromatogr. B.* 1998; 718: 87.
50. Ayrton J, Dear GJ, Leavens WJ, Mallett DN, Plumb RS. *Rapid Commun. Mass Spectrom.* 1997; 11: 1953.
51. Beaudry F, LeBlanc JYC, Coutu M, Brown N. *Rapid Commun. Mass Spectrom.* 1998; 12:1216.
52. Sen JW, Bergen HR 3rd, Heegaard NH. *Anal Chem.* 2003; 75: 1196.
53. Wang S, Mei H, Ng K, Workowski K, Astle T, Korfmacher W. *Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, May 26–30, 2002.
54. Abola E, Kuhn P, Earnest T, Stevens RC. *Nat. Struct. Biol.* 2000; 7: 973.
55. Kuhn P, Wilson K, Patch MG, Stevens RC. *Curr. Opin. Chem. Biol.* 2002; 6:704.
56. Stout TJ, Foster PG, Matthews DJ. *Curr. Pharm. Des.* 2004; 10:1069.
57. Keifer PA. *Curr Opin Chem Biol.* 2003; 7: 388.
58. Taylor LCE, Johnson RL, Raso R. *J. Am Soc Mass Spectrom.* 1995; 6: 387.
59. Greaves J. *J Mass Spectrom.* 2002; 8: 777.
60. Mallis LM, Sarkahian AB, Kulishoff JM Jr, Watts WL Jr. *J Mass Spectrom.* 2001; 9: 889.
61. Morand KL, Burt TM, Regg BT, Chester TL. *Anal. Chem.* 2001; 73: 247.
62. Romanyshyn L, Tiller PR, Alvaro R, Pereira A, Hop CECA. *Rapid Commun. Mass Spectrom.* 2001; 15: 313.
63. Wu J-T, Zeng H, Deng Y, Unger SE. *Rapid Commun. Mass Spectrom.* 2001; 15: 1113.
64. King RC, Miller-Stein C, Magiera DJ, Brann J. *Rapid Commun. Mass Spectrom.* 2002; 16: 43.
65. Fang L, Cournoyer J, Demee M, Zhao J, Tokushige D, Yan B. *Rapid Commun Mass Spectrom.* 2002; 16: 1440.
66. Xu R, Wang T, Isbell J, Cai Z, Sykes C, Brailsford A, Kassel DB. *Anal Chem.* 2002; 74: 3055.
67. Yang L, Mann TD, Little D, Wu N, Clement RP, Rudewicz PJ. *Anal Chem.* 2001; 73: 1740.
68. Khandurina J, Guttman A. *J Chromatogr A.* 2002; 943: 159.
69. Weigl BH, Bardell RL, Cabrera CR. *Adv Drug Deliv Rev.* 2003; 55: 349.
70. Huikko K, Kostiainen R, Kotiaho T. *Eur J Pharm Sci.* 2003; 20: 149.
71. Jakeway SC, de Mello AJ, Russell EL. *Fresenius J Anal Chem.* 2000; 366: 525.
72. Schultz GA, Corso TN, Prosser SJ, Zhang S. *Anal. Chem.* 2000; 72: 4058.
73. Meng F, Du Y, Miller LM, Patrie SM, Robinson DE, Kelleher NL. *Anal. Chem.* 2004; 76: 2852.
74. Keetch CA, Hernanndez H, Sterling A, Baumert M, Allen MH, Robinson CV. *Anal Chem.* 2003; 75: 4937.
75. Dethy JM, Ackermann BL, Delatour C, Henion JD, Schultz GA. *Anal. Chem.* 2003; 75: 805.
76. *Handbook of Pharmaceutical Salts Properties, Selection, and Use* P. Heinrich Stahl, Camille G. Wermuth (Eds.) 2002.

BIBLIOGRAPHY

Mass Spectrometry

1. Busch KL, Glish GL, McLuckey SA. *Mass spectrometry/mass spectrometry,* VHC, New York, 1988.
2. Watson JT. *Introduction to Mass Spectrometry*, 3rd ed., Lippincott-raven, Philadelphia, 1997.
3. Willoughby R, Sheehan E, Mitrovich S. *A Global view of LC/MS*, Global View Publishing, Pittsburgh, 1998.
4. The American Society of Mass Spectrometry web page and associated links. (WWW.asms.org).
5. Murray KK. *J. Mass Spectrom.* 1999; 34: 1.

NMR

1. Skoog DA. *Principles of Instrumental Analysis*. Saunders College Publishing, Philadelphia, 1984
2. Lambert JB, Shurvell HF, Lightner DA, Cooks RG. *Organic Structural Spectroscopy*. Prentice Hall, Upper Saddle River, NJ, 1998.
3. Keifer PA. In ed. Jucker E. *In Progress in Drug Research*, Vol 55, Birkhauser Verlag, Basel, Switzerland, 2000.

Instrumental Methods of Analysis, General

1. Analytical Chemistry, Fundamental Reviews. Washington, DC: American Chemical Society, April 1982.
2. Borman SA. *Instrumentation in Analytical Chemistry, vol 2*. Washington, DC: American Chemical Society, 1982.
3. Christian GD, O'Reilly JE. *Instrumental Analysis*, 2nd ed. Boston: Allyn & Bacon, 1986.
4. Ewing GW. *Instrumental Methods of Chemical Analysis,* 5th Ed. New York, McGraw-Hill, 1985.
5. Mann CK. *Instrumental Analysis*. New York: Harper & Row, 1974.
6. Moore WJ. *Physical Chemistry*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 1972.
7. Munson JW. *Pharmaceutical Analysis, Modern Methods*, Parts A, B. New York: Dekker, 1981, 1984.
8. Schirmer RE. *Modern Methods of Pharmaceutical Analysis*, 2nd Ed, vols 1, 2. Boca Raton, FL: CRC Press, 1991.
9. Willard HH. *Instrumental Methods of Analysis*, 6th Ed. New York: Van Nostrand, 1981.

Ultraviolet and Visible Spectrometry

1. *ASTM Index to Ultraviolet and Visible Spectra*. ASTM Tech Publ 357. Philadelphia: ASTM, 1963.
2. Braude EA. *Determination of Organic Structures by Physical Methods*. New York: Academic, 1955, pp 131–194.
3. Duncan ABF, Matsen FA. In: Weissberger A, ed. *Technique of Organic Chemistry*, vol 9, 2nd Ed. New York: Interscience, 1968–1970, p 581.
4. Harris TD. Anal Chem 1982; 54: 741A.
5. Hershenson HM. *Ultraviolet and Visible Absorption Spectra*, Index, 1930–1963, 6 vols. New York: Academic, 1966.
6. Jaffe HH, Orchin M. *Theory and Applications of Ultraviolet Spectroscopy*. New York: Wiley, 1962.
7. Lang L. ed. *Absorption Spectra in the Ultraviolet and Visible Region*, vols 1–17. New York: Academic, 1961–1973.
8. Montegu B, Langier A. *Fournier J. J Phys* [E] 1979; 12: 1153.
9. Organic Electronic Spectral Data, 1946–1967, vols 1–7. New York: Interscience, 1960–1971.
10. Scott AI. *Interpretation of the Ultraviolet Spectra of Natural Products*. New York: Pergamon, 1964.

Infrared Spectrometry

1. Bellamy LJ. *The Infrared Spectra of Complex Molecules*, 3rd Ed. New York: Wiley, 1975.
2. Colthup NB. *Introduction to Infrared and Raman Spectroscopy*, 3rd Ed. New York: Academic, 1990.
3. Dyer JR. *Organic Spectral Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1972.
4. Griffiths PR, de Haseth JA. *Fourier Transform Infrared Spectrometry*. New York: Wiley, 1986.
5. Hershenson HM. *Infrared Absorption Spectra*, Index, 1947–1954, 2 vols. New York: Academic, 1965.
6. Hurley WJ. *J Chem Educ* 1966; 43: 236.
7. Lang L. ed. *Absorption Spectra in the Infrared Region*, vols 1, 2. London: Butterworths, 1974, 1976.
8. Low MJD. *J Chem Educ* 1970; 47: A163, A255, A415.
9. Fourier MA. *Hadamard and Hilbert Transforms in Chemistry*. New York: Plenum, 1982.
10. Martin AE. *Infrared Instrumentation and Techniques*. New York: Elsevier, 1966.
11. Massart DL, Vandefinste BGM, Deming SN. *Chemometrics: A Textbook*. New York: Elsevier, 1988.
12. Nyquist RA, Kegel RO. *Infrared Spectra of Inorganic Compounds*. New York: Academic, 1971.
13. *Catalog of Infrared Spectra. Philadelphia*: Sadtler Research Labs, nd.
14. Silverstein RM et al. *Spectrometric Identification of Organic Compounds*, 5th Ed. New York: Wiley, 1991.
15. Szymanski HA, Erickson RE. *Infrared Band Handbook*, rev Ed. New York: Plenum, 1970. Suppls 1 and 2 cover the 200 to 600 cm^{-1} region.
16. Vornhederand PF, Brabbs WJ. *Anal Chem* 1970; 42: 1454.

Emission Spectrometry, Flame Photometry, and Atomic Absorption Spectrometry

1. Ahrens LH, Taylor SR. *Spectrochemical Analysis*, 2nd ed. Reading, MA: Addison-Wesley, 1961.
2. Alkemade CTJ, Milatz JMW. *Appl Sci Res* 1955; B4: 289.
3. Alkemade CTJ, Milatz JMW. *J Opt Soc Am* 1955; 45: 583.
4. Brode WR. *Chemical Spectroscopy*, 2nd ed. New York: Wiley, 1943.
5. Dedina J, Rubeska I. *Spectrochim Acta B* 1980; 35B: 119.

6. Elwell WT, Gidley JAF. *Atomic Absorption Spectrophotometry*, 2nd rev Ed. New York: Pergamon, 1966.
7. Godden RG, Thomerson DR. *Analyst* 1980; 105: 1137.
8. Haswell SJ, ed. *Atomic Absorption Spectrometry*. New York: Elsevier, 1991.
9. Mavrodineanu R, ed. *Analytical Flame Spectroscopy*. Berlin: Springer-Verlag, 1971.
10. Meggers WF, et al. *Tables of Spectral-Line Intensities, parts 1, 2*. National Bureau of Standards (US) Monograph 32. Washington, DC: USGPO, 1961–1962. Revised edition: Corliess CH, 1967.
11. Pinta M. *Atomic Absorption Spectrometry*, vol 2. Application to Chemical Analysis, 2nd ed. Paris: Masson, 1980.
12. Reed TB. *J Appl Phys* 1961; 32: 821, 2534.
13. Reed TB. *Int Sci Technol*. June 1962; 142.
14. Styris DL, Kaye JH. *Spectrochim Acta B* 1981; 36B: 41.
15. Van Loon JC. *Analytical Atomic Absorption Spectroscopy, Selected Methods*. New York: Academic, 1980.
16. Walsh A. *Spectrochim Acta* 1955; 7: 108.
17. Willard H, et al. *Instrumental Methods of Analysis*, 6th Ed. New York: Van Nostrand, 1981.

Fluorescence and Phosphorescence Spectrometry

1. Guilbault GC. *Fluorescence: Theory, Instrumentation and Practice*. New York: Dekker, 1967.
2. Guilbault GC, ed. *Practical Fluorescence*. New York: Dekker, 1990.
3. Hercules DM, ed. *Fluorescence and Phosphorescence Analysis: Principles and Applications*. New York: Interscience, 1966.
4. Udenfriend S. *Fluorescence Assay in Biology and Medicine*. New York: Academic, 1962.

Light Scattering and Polarimetry

1. Crabbe P. *ORD and CD in Chemistry and Biochemistry*. New York: Academic, 1972.
2. Djerassi C. *Optical Rotatory Dispersion*. New York: McGraw-Hill, 1960.
3. Stacey K. *Light-Scattering in Physical Chemistry*. London: Butterworths, 1956.
4. Weissberger A, ed. *Physical Methods in Organic Chemistry*, vol 1, 3rd Ed, Part 2. New York: Interscience, 1960.

Dissolution

Vijai Kumar, MS, MBA

Divya Tewari, M Pharm, MS

Dissolution is the process by which a solid enters into solution. The earliest reference to dissolution is probably the 1897 article by Noyes and Whitney, titled as "The Rate of Solution of Solid Substances in Their Own Solution." The authors suggested that the rate of dissolution of solid substances is determined by the rate of diffusion of a very thin layer of saturated solution that forms instantaneously around the solid particle. They developed the mathematical relationship that correlates the dissolution rate to the solubility gradient of the solid. Their equation is still the basic formula upon which most of the modern mathematical treatments of the dissolution phenomenon revolve.

Interestingly, the work of Noyes and Whitney, together with the studies that followed in the early part of the 20th century, was primarily based on the physicochemical aspects of dissolution applied to chemical substances. The most prominent part of these investigations that deserve recognition are those of Nernst and Brunner in 1904 for their application of Fick's law of diffusion to the Noyes-Whitney equation, and those of Hixson and Crowell in 1931 for their development of famous "Cube Root Law" of dissolution.[1]

By the middle of the 20th century, emphasis started to shift to the examination of the effects of dissolution behavior of drugs on the biological activity of pharmaceutical dosage forms. One of the earliest studies with this purpose in mind was conducted by J Edwards in 1951 on aspirin tablets. He reported, "because of its poor solubility, the analgesic action of aspirin tablets would be controlled by its dissolution rate within the stomach and the intestine." No *in vivo* studies, however, were conducted by Edwards to support his postulate.

About 8 years later, Shenoy and colleagues proved the validity of Edward's suggestion of the *in vitro/in vivo* correlation by demonstrating a direct relationship between the bioavailability of amphetamine from sustained-release tablets and its *in vitro* dissolution rate. Other studies, especially those reported by Nelson, Levy, and others, confirmed beyond doubt the significant effect of the dissolution behavior of drugs on their pharmacological activities. Because of the importance of these findings, dissolution testing began to emerge as a dominant topic within both the pharmaceutical academia and the drug industry.

In the late 1960s dissolution testing became a mandatory requirement for several dosage forms. The role of dissolution in the absorption of drug products, however, still is far from being understood completely. In spite of the reported success of several *in vitro/in vivo* correlation studies, dissolution cannot be relied upon as a predictor of therapeutic efficiency. Rather, it is a qualitative tool that can provide valuable information about the biological availability of a drug as well as batch-to-batch consistency. Another area of difficulty is the fact that the accuracy and precision of the testing procedure is dependent, to a large extent, on the strict observance of so many subtle parameters and detailed operational controls.

In spite of these shortcomings, dissolution is considered today as one of the most important quality control procedure performed on pharmaceutical dosage forms. Whether or not it has been correlated with biological effectiveness, the standard dissolution test is a simple and inexpensive indicator of product's physical consistency. If one batch differs from the other in its dissolution characteristics, or if the dissolution profiles of the production batches show a consistent trend upwards or downwards, it sounds a sure warning that some factor in the raw material, formulation, or process is out of control.[1] Additionally, dissolution data seems to be a useful tool in the early stages of drug development and molecular manipulation. In the early stages of research, steps may be taken to optimize characteristics that will influence subsequent data concerning biological availability. Based on simple dissolution test, selection of a proper salt for a new drug can be done at early drug development stage.

Definition of Dissolution and Theoretical Concepts for the Release of the Drug from Dosage Forms

"Dissolution is defined as the process by which solid substances enters in solvent to yield a solution. Stated simply, dissolution is the process by which a solid substance dissolves. Fundamentally, it is controlled by the affinity between the solid substance and the solvent." The physical characteristics of the dosage form, the wettability of the dosage unit, the penetration ability of the dissolution medium, the swelling process, the disintegration and the deaggregation of the dosage forms are few of the factors that influence the dissolution characteristics of drugs. Wagner proposed a scheme depicted in Figure 35-1 for the processes involved in the dissolution of solid dosage forms.

This scheme was later modified to incorporate other factors that precede the dissolution process of solid dosage forms. Carstensen proposed a scheme incorporating the following sequence:

1. Initial mechanical lag
2. Wetting of the dosage form
3. Penetration of the dissolution medium into the dosage form
4. Disintegration
5. Deaggregation of the dosage form and dislodgement of the granules
6. Dissolution
7. Occlusion of some particles of the drug

Carstensen explained that the wetting of the solid dosage form surface controls the liquid access to the solid surface and, many times, is the limiting factor in the dissolution process. The speed of wetting directly depends on the surface tension at the interface (interfacial tension) and upon the contact angle, θ, between
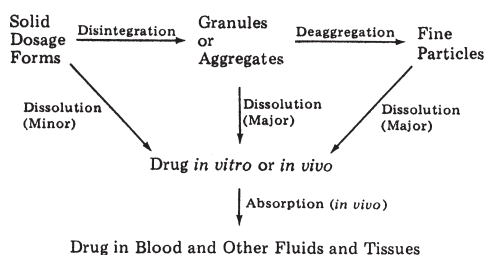
Figure 35-1. Dissolution process of solid dosage forms.

the solid surface and the liquid. Generally, a contact angle of more than 90° indicates poor wettability. Incorporation of a surfactant, either in the formulation or in the dissolution medium, lowers the contact angle and enhances dissolution. Also, the presence of air in the dissolution medium causes the air bubbles to be entrapped in the tablet pores and act as a barrier at the interface. For capsules, the gelatin shell is extremely hydrophilic, and therefore, no problems in wettability exist for the dosage itself (although it may exist for the powders inside).

After the solid dosage form disintegrates into granules or aggregates, penetration characteristics play a prime role in the deaggregation process. Hydrophobic lubricants, such as talc and magnesium stearate, commonly employed in tablet and capsule formulations, slow the penetration rate and, hence, the deaggregation process. A large pore size facilitates penetration, but if it is too large it may inhibit penetration by decreasing the internal strain caused by the swelling of the disintegrant.

After deaggregation and dislodgment occur, the drug particles become exposed to the dissolution medium and dissolution proceeds as previously discussed under Film Theory. Figure 35-2[2] graphically presents the model proposed by Carstensen.

It is apparent from Figure 35-2 that the rate of dissolution of the drug can become rate-limiting step before it appears in the blood. However, when the dosage form is placed into the gastrointestinal tract in solid form, there are two possibilities for the rate-limiting step. The solid must first dissolve, and the drug in solution must then pass through the gastrointestinal (GI) membrane. Freely water-soluble drugs will tend to dissolve rapidly, making the passive diffusion of the drug or the active transport of the drug rate-limiting step for absorption through the GI membrane. Conversely, the rate of absorption of poorly water-soluble drugs will be limited by the rate of dissolution of the undissolved drug or disintegration of dosage form.

The rate of dissolution of drug substance is determined by the rate at which solvent-solute forces of attraction overcome the cohesive forces present in the solid. This process is rate-limiting when the release of solute into solution is slow and the transport into the bulk solution is fast. In t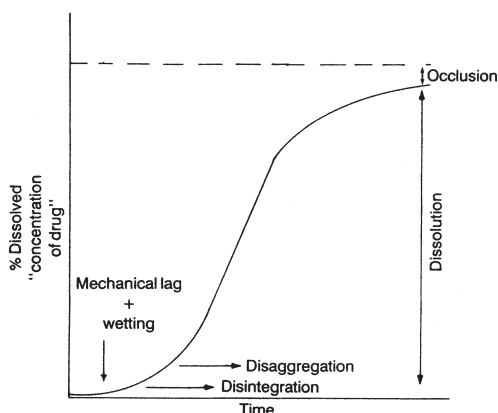his case the dissolution is said to be interfacially controlled. Dissolution may also be diffusion controlled, where the solvent-solute interaction is fast compared to transport of solute into the bulk solution. In diffusion-controlled process, a stationary layer of solute adjacent to the solid/ liquid interface is postulated and is commonly referred to as the diffusion layer. The saturation concentration of solute develops at the interface and decreases with distance across the diffusion layer.

MATHEMATICS OF DISSOLUTION

It has long being recognized that the release of the active drug from a drug product may be greatly influenced by the physicochemical properties of the drug as well as the dosage form.[3] The availability of the drug is usually determined by the rate of release of the drug from the physical system (dosage form). The release of the drug from its dosage form is usually determined by the rate at which it dissolves in the surrounding medium. The rate of dissolution of a chemical or drug from the solid state is defined as the amount of drug substance that goes into solution per unit time under standardized condition of liquid/solid interface, temperature, and solvent composition. In biopharmaceutics, rate of dissolution usually refers to the rate at which the drug dissolves from an intact dosage form or from fragments or particles from the dosage form during the test.[4]

The following section deals with the introductory concepts on mathematics of dissolution focusing primarily on intrinsic dissolution.

Intrinsic Dissolution

The rate of dissolution of a pure pharmaceutical active ingredient when conditions such as surface area, temperature, agitation or stirring speed, pH, ionic strength of the dissolution medium is kept constant is known as intrinsic dissolution rate. This parameter allows the screening of the drug candidates and aids in understanding their solution behavior under various biophysiological conditions.[5]

INTRINSIC DISSOLUTION RATE CONSTANTS—The rate at which a substance dissolves in a liquid to form a solution is governed by physical parameters such as the surface area of the substance at a given time during the process of dissolution, the shape of the substance, the characteristics of the solid/ liquid interface, and the solubility of the substance in the liquid. Hence, dissolution can be considered a specific type of certain heterogeneous reaction which results in a mass transfer as a net effect between the escape and deposition of solute molecules at a solid surface. Mathematically, the process can be simply described as follows:

$$dM/dT = KA(C_s - C) \qquad (1)$$

where, M is the mass of the substance remaining to be dissolved, A is the surface area exposed to the dissolution medium, Cs is the saturation concentration referred to as solubility in the dissolution medium, C is the amount dissolved or the concentration of the drug in solution at time t, K is the intrinsic dissolution rate constant or simply the dissolution rate constant.

The equation expresses the fact that when C is small, $C < 0.15C_s$, then K is proportional to C_s, since $(C_s - C)$ is large. If this applies, then to a good approximation we may write

$$dM/dT = KAC_s \qquad (2)$$

Equation 2 is commonly referred to as a sink-condition equation, which implies that sink conditions exist during the process of dissolution. It must be noted, however, that A is a constant except initially, when only very small quantities of solute have dissolved and where there is an amount of solute far in excess of saturation.

Figure 35-2. The S-shaped dissolution curve of solid dosage forms.

When the process of dissolution takes place under sink conditions, a stagnant film of liquid (dissolution medium) is adsorbed onto the solid, the thickness of this film being l cm. The liquid in the film that is in direct contact with the solid is saturated with drug in solution. The concentration of the drug in solution then drops as the distance from the dissolving solid surface increases. At the end of the film, l cm from the surface, the concentration in the film is the same as that in the bulk solution, C_b. The driving force behind the movement of solute molecules through the stagnant film is the concentration gradient that exists between the saturation concentration of the solute, C_s, in the stagnant layer at the surface of the solid and its concentration on the farthest side of the stagnant film, C_b. A schematic representation of dissolution as a physicochemical phenomenon is shown in Figure 35-3, the greater this difference in concentration, the faster the rate of dissolution.

Applying Fick's first law of diffusion to equation 2, the flux, J (defined as the rate of flow of material through 1 cm^2, ie,

$$J = \frac{dM/dT}{A}$$

can be expressed as

$$J = -D(\delta C/\delta x)$$

where D is the coefficient of diffusion; x is the distance as shown in the Figure 35-3. If the concentration gradient, $\delta C/\delta x$, is linear, if $C = C_s$ at the surface ($x = 0$), and if $C = C_b$ (the bulk concentration at the interface between the bulk solution and the film, where $x = 1$), then

$$(\delta C/\delta x) = (C_b - C_s/l) \tag{4}$$

Therefore,

$$\frac{1}{A}\frac{dM}{dT} = -D\frac{C_b - C_s}{l} = -K(C_s - C)$$
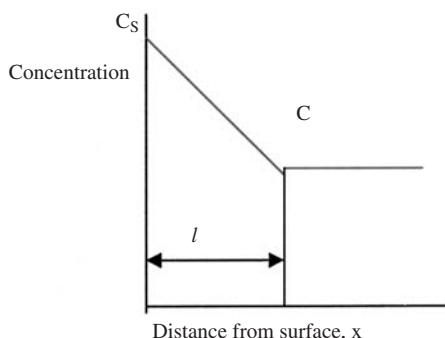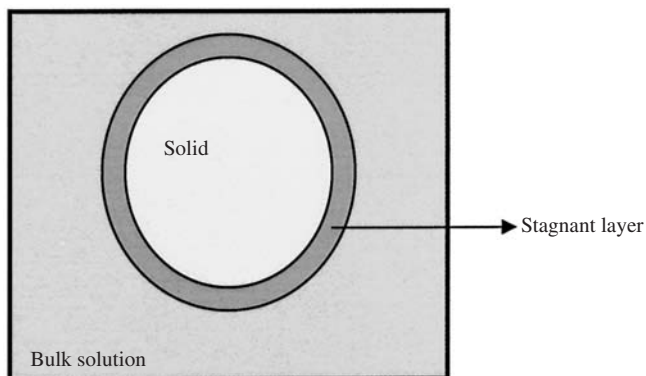
Figure 35-3. Physical model depicting a dissolution process.

Or simply

$$\frac{dM}{dT} = -KA(C_s - C)$$

If the agitation intensity of the system containing suspended particles is increased, the thickness of the film will decrease progressively. Hence k is a function of the test as well. Additionally, if the product of A (Cs 2 C) is maintained constant for many drug tested by the same test, the relative magnitude of k values will indicate the effective ease of dissolution. In practice, k also includes the dynamics of the shear rate between solid and the solvent, that is, the rate at which fresh solvent contacts the surface of the solid; highly complex processes, including diffusion rate through the boundary layers, depend upon this rate.

The shear rate depends upon a multitude of variables that must be controlled if the test is to be repeatable. Those variables include the flow pattern of solvent in the apparatus, turbulence, viscosity, surface tension, and dissolved gases, all of which are subject to uncontrolled input variables to the system, such as vibrations and system geometry. The theoretical basis for those inputs rests in the realm of chemical engineering's fluid flow and surface boundary theories, which are discussed elsewhere.[6]

The intrinsic dissolution rate has been used as a means to demonstrate the chemical purity and equivalency of the active pharmaceutical ingredient (API). The use of rotating disk system (USP Wood Apparatus), which is similar to USP procedure 1 is most common, though stationary disk systems, vertical diffusion cells, and enhancer cells can also be used to measure the intrinsic dissolution rates.[7]

Rotating Disk System (USP Wood Apparatus)

The apparatus consists of steel punch, a die, and a base plate. The base of the die has three threaded holes for the insertion of the screws and the attachment to the base plate. The material is placed in the die cavity, and the punch is inserted into the cavity and the material is compressed. The die is screwed onto the shaft holder, and the shaft holder is mounted on the stirring device. The shaft is a stainless steel rod with hollow die holder. Pellet and the die assembly are introduced all at once, when the dissolution drive mechanism is lowered. The dissolution is achieved by shear-like motion of the pellet in the dissolution medium. The dissolution vessel is standard curved bottom 1-L flask. Care should be taken that the air bubbles do not form on the surface of the pellet or else it will interfere with the dissolution rate (Fig 35-4).

Stationary Disk System

The apparatus consists of a steel punch, a die, and a base plate. The die base has three holes for the attachment of the base plate. The three fixed screws on the base plate are inserted through the three holes on the die. Punch is inserted into the die cavity filled with the material, and the material is then compressed. The pellet and die assembly can then be inserted with the pellet side up, into the bottom of the dissolution vessel, which is flat bottomed. The USP Apparatus 2 is the stirring mechanism here. The advantage of this system is that no air bubbles are formed on the pellet surface. There is also no change in the temperature as the device is small and is totally submerged into the dissolution medium (Fig 35-5).

Correlation Between the Disintegration and Dissolution

The close correlation between disintegration and dissolution has been studied by many investigators. Both processes exhibit
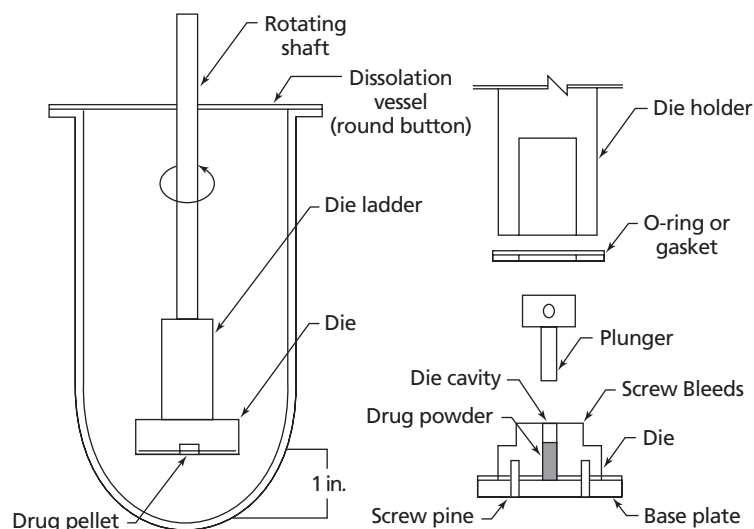
Figure 35-4. Rotating disk system (USP Wood apparatus).

"S"-shaped curves and a probit or a weibul function was suggested to explain the data. In general, however, disintegration has proved to be a poor indicator of bioavailability because of the turbulent agitation maintained during the test. Several other factors such as solubility, particle size, and crystalline structure, among others, have been found to affect the dissolution of the drug substance but have no relevance to disintegration.

Factors Affecting the Rate of Dissolution

The dissolution rate data can be meaningful only if the results of successive test on the same dosage form are consistent within reason. The dissolution test should yield reproducible result even when it is performed in different laboratories or with different personnel. To achieve high reproducibility, all variables that influence the test should be clearly understood and possibly controlled.

Factors affecting the dissolution rate of drugs from a dosage form include the following:

1. Factors related to the physicochemical properties of the drug
2. Factors related to drug product formulation
3. Effect of processing factors on the dissolution rate
4. Factors related to dissolution test parameters
5. Miscellaneous factors

FACTORS RELATED TO THE PHYSICOCHEMICAL PROPERTIES OF THE DRUG

EFFECT OF SOLUBILITY ON DISSOLUTION—The physicochemical properties of the drug substance play a prime role in controlling its dissolution from the dosage form. The modified Noyes and Whitney equation shows that the aqueous solubility of the drug is the major factor that determines its dissolution rate. Actually, some studies showed that drug-solubility data could be used as a rough predictor of the possibility of any future problems with bioavailability, a factor that should be taken into consideration in the formulation design.

EFFECT OF PARTICLE SIZE ON DISSOLUTION— According to Nernst- Brunner theory, the dissolution rate is directly proportional to the surface area of the drug. Since the surface area increases with the decreasing particle size, higher dissolution rates may be achieved through the reduction of the particle size. This effect has been highlighted by the superior dissolution rate observed after "micronization" of certain sparingly soluble drugs as opposed to the regularly milled form.

Several investigations have demonstrated an increased absorption rate for griseofulvin after micronization. Similar effects have been reported for chloramphenicol, tetracycline salts, sulfadiazine, and norethisterone acetate. In the case of chloramphenicol, it has been shown that formulations containing smaller particles (50–200 μm) were absorbed faster than formulations containing larger particles (400–800 μm). Figure 35-6 presents the effect of particle-size differences on the dissolution rate of phenacetin and phenobarbital.[8]

However, when employing this technique to enhance dissolution, it is important to recognize the fact that it is the effective surface area that has to be increased. The effective surface area is the surface area available to the dissolution fluid. If the drug is hydrophobic and the dissolution medium has poor wetting properties, reduction of particle size may lead to decreased effective surface area and hence a "slower" rate of dissolution.

Physical properties of the drug particles other than size also affect indirectly the effective surface area by modifying the shear rate of the fresh solvent that comes in contact with the solid. These properties include the particle shape and the density.

The mechanism by which the reduction in particle size improves dissolution is usually through the enhancement of the drug solubility. It is assumed that the drug solubility is independent of particle size. However, the drug solubility and the
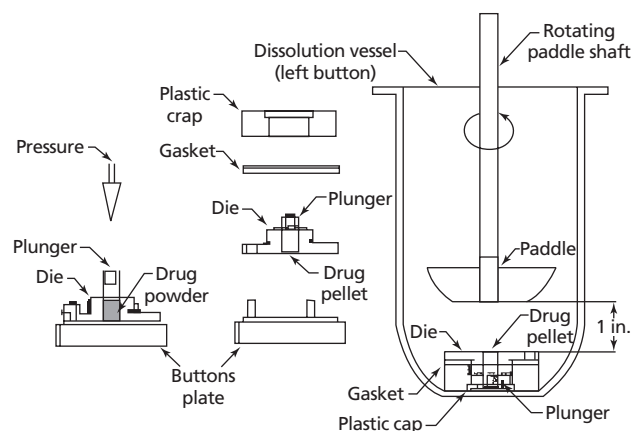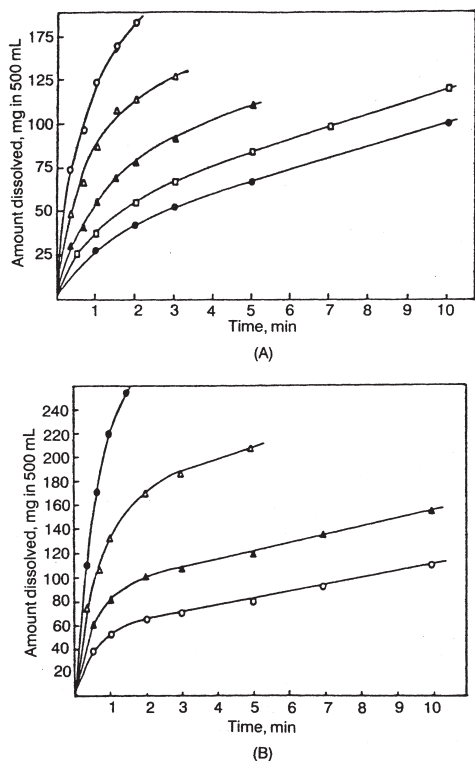

Figure 35-5. Stationary disk system (new apparatus).

Figure 35-6. Effect of particle size on the dissolution rate of drugs from solid dosage forms.[3] **A.** Phenacetin: ○ particle size: 0.11–0.15 mm; △ particle size: 0.15–0.21 mm; ▲ particle size: 0.21–0.30 mm; □ particle size: 0.30–0.50 mm; ● particle size: 0.50–0.71 mm. **B.** Phenobarbital ● particle size: 0.07–0.15 mm; △ particle size: 0.15–0.25 mm; ▲ particle size: 0.25–0.42 mm; ○ particle size: 0.42–0.71 mm.

surface area can be correlated by the *Ostwald-Freundlich equation:*

$$\ln S = \frac{2M\gamma}{\rho RT} \frac{1}{r} = \frac{\alpha}{r}$$

where, M is the molecular weight, ρ is the density, γ is the interfacial tension or surface free energy of the solid, T is the temperature, R is the gas constant, r is the radius of the particle.

From the above equation

$$S = S_\infty \cdot e^{\frac{\alpha}{r}}$$

The equation shows that the solubility is inversely proportional to particle radius. Therefore, *S* could be viewed as the solubility of the microparticles and S_∞ as the solubility of the macro particles. However, it is obvious that the particle radius has to be reduced to a microlevel before it can effect a change in solubility.

This extreme reduction in particle size usually cannot be achieved through regular milling or even micronization procedures, and therefore other methods have been recommended. One of these involves formation of a *solid solution* or *molecular dispersion* where the molecules of the sparingly soluble drug either are dispersed interstitially in a water-soluble drug or replaced in its crystal lattice.

Another technique, which also produces extremely small particles but still larger than the ones produced by solid solution, is by dispersion of the drug into a soluble carrier such as polyvinylpyrrolidone (PVP) solution. These techniques usually are employed for the enhancement of dissolution rate of insoluble drugs.

EFFECT OF SOLID PHASE CHARACTERISTICS OF THE DRUG ON DISSOLUTION—Amorphicity and crystallinity, the two important solid-phase characteristics of drugs affect their dissolution profile. Numerous studies have demonstrated that the amorphous form of a drug usually exhibits greater solubility and higher dissolution rate as compared to that exhibited by the crystalline form. For example, it was shown that the amorphous form of novobiocin has a greater solubility and higher dissolution rate than the crystalline form. Blood-level studies confirmed such findings where administration of the amorphous form yielded about three to four times the concentration compared to the administration of the crystalline form. Similar differences were demonstrated for griseofulvin, phenobarbital, cortisone acetate, and chloramphenicol. Chloramphenicol palmitate is one example that exists in at least two polymorphs. The B form is apparently more bioavailable. The recommendation might be that manufacturers should use polymorph B for maximum absorption. One contradictory example is that of erythromycin esteolate, where the dissolution rate of amorphous form is markedly lower than the crystalline form of erythromycin esteolate, as exemplified by Figure 35-7. However, a method of controlling and determining crystal form would be necessary in the quality control process.[9]

EFFECT OF POLYMORPHISM ON DISSOLUTION—Polymorphic forms of drugs have been shown to influence changes in solubilizing characteristics and thus the dissolution rate of the drug in question. Numerous reports have shown that polymorphism and the state of hydration, solvation, and/or complexation markedly influence the dissolution characteristics of the drug. The drugs that exhibit influence on the dissolution behavior include tolbutamide, chloramphenicol, and others.

FACTORS RELATED TO DRUG PRODUCT FORMULATION

It has been shown that the dissolution rate of a pure drug can be altered significantly when mixed with various excipients during the manufacturing process of solid dosage forms. These excipients are added to satisfy certain pharmaceutical functions such as diluents (fillers), dyes, binders, granulating agents, disintegrants, and lubricants. Generically identical tablet and capsule products, manufactured by different pharmaceutical manufacturers, were found to exhibit significant differences in dissolution rates for their active ingredients. In certain cases, several studies showed that poor tablet and capsule formulations have been shown to cause a marked decrease in bioavailability and impairment of the clinical response. Such findings during the 1960s, especially in the case of digoxin and tolbutamide tablets, as well as chloramphenicol and tetracycline HCl (all lifesaving drugs), were the triggering factors that compelled the drug-regulatory agencies and compendial au-
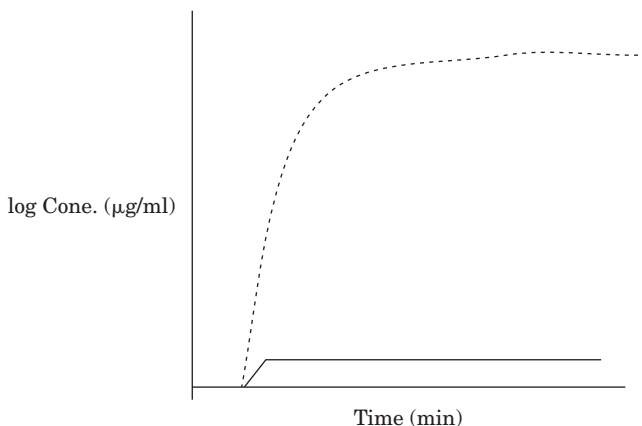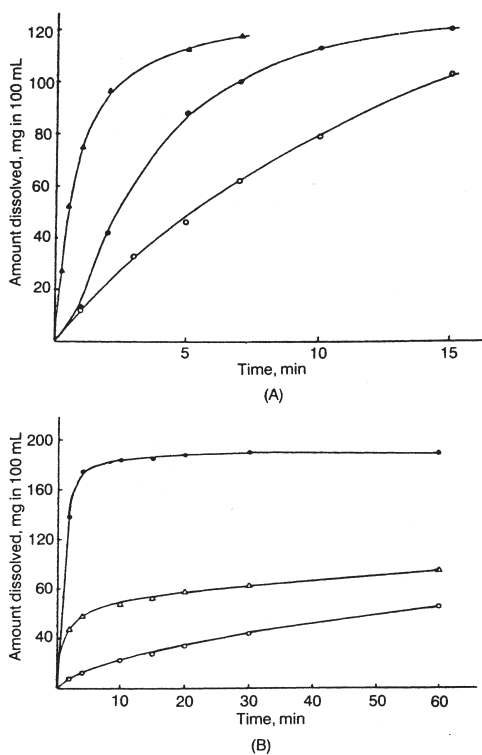

Figure 35-7. Dissolution performance of erythromycin estolate. (dotted line), Crystalline form; (solid line), Amorphous form.

thorities to institute the dissolution test as a legal requirement for most solid dosage forms.

EFFECT OF GRANULATING AGENTS AND BINDERS—Solvang and Finholt [10] have shown that Phenobarbital tablets granulated with gelatin solution provide faster dissolution rate in gastric fluid than those prepared using sodium carboxymethylcellulose or polyethylene glycol 6000 as a binder. This observation was attributed to the fact that gelatin imparts hydrophilic characteristics to the hydrophobic drug surface, whereas PEG 6000 forms complex with poor solubility, and sodium carboxymethylcellulose is converted to its less soluble acid form at low pH of the gastric fluid (Figure 35-8). Even gelatin obtained from various processes and origins has been shown to affect the dissolution rate of dosage forms. [11]

Various studies have been reported in the literature evaluating the effects of various granulating agents and binders on the dissolution rate of tablets. [12]

EFFECT OF DISINTEGRANTS AND DILUENTS—The type and amount of disintegrating agent employed in the formulation significantly controls the overall rate of dissolution of dosage form. Jaminet et al [11] employed several disintegrating agents in manufacturing of Phenobarbital tablets, including Primojel (sodium glycolate of potato starch), Nymcel (polymerized water-soluble brand of sodium carboxymethylcellulose), and Copagel (low viscosity grade of sodium carboxymethylcellulose). The effect on the dissolution rate of tablets by the addition of disintegrants before and after granulation was assessed. When added before granulation, Copagel gave tablets with a remarkably slow dissolution rate. However, when added after granulation, Copagel did not result in lowering the dissolution rate. Primojel was not found to be as effective, particularly on addition after granulati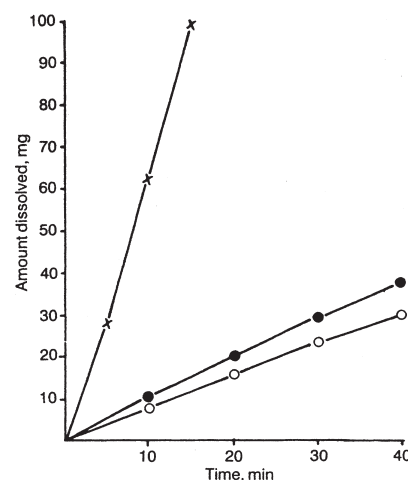on. Levy, in 1963, studied the effect of starch, the most commonly used diluent, on the rate of dissolution of salicylic acid tablets manufactured by the dry, double-compression process [13] (Fig 35-9). Increasing the starch content from 5% to 20% resulted in a dramatic increase in the dissolution rate (almost threefold). This was attributed to better and more thorough disintegration. Later, however, Finholt suggested that the hydrophobic drug crystals acquire a surface layer of fine starch particles that imparts a hydrophilic property to the granular formulation and thereby increases the effective surface area and hence the dissolution rate (see Fig 35-9).

EFFECT OF LUBRICANTS—The nature, quality, and quantity of lubricants added can affect the dissolution rate. The effect of various lubricants on dissolution rate of salicylic acid was studied and it was concluded that magnesium stearate, a hydrophobic lubricant, tends to retard the dissolution rate of salicylic acid tablets, whereas sodium lauryl sulfate enhances dissolution, due to its hydrophobic character combined with surface activity, which increases the microenvironment pH surrounding the weak acid and increases wetting and better solvent penetration into the tablets [14] (Fig 35-10) illustrates the effect of lubricants on the dissolution rate of tablets.
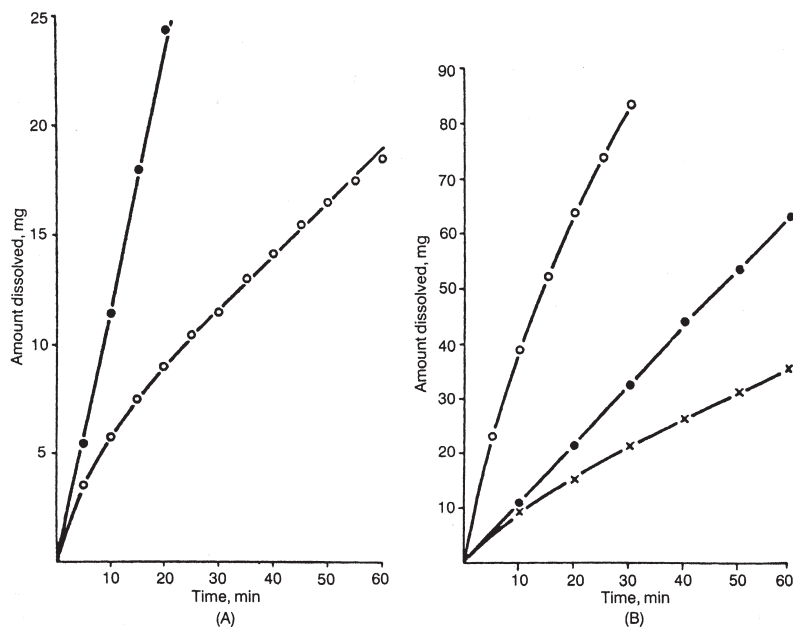
Effect of lubricants on the dissolution rate of drugs from dosage form would depend on properties of the granules, the lubricant itself, and the amount of lubricant used. If granules are hydrophilic and fast disintegrating, a water-soluble surface-active lubricant will have an insignificant effect on the dissolution. On the other hand, if the granules are hydrophobic, the surface-active lubricant will enhance dissolution. It was also found that hydrophobic lubricants, such as magnesium stearate, aluminum stearate, stearic acid, and talc, decrease the effective drug-solvent interfacial area by changing the surface characteristics of the tablets, which results in reducing its wettability, prolonging its disintegration time, and decreasing the area of the interface between the active ingredient and solvent.

FACTORS RELATED TO THE DISSOLUTION TEST PARAMETERS

METHOD OF GRANULATION—Wet granulation has been shown to improve the dissolution rates of poorly soluble drugs by imparting hydrophilic properties to the surface of the granules. Additionally the use of fillers and diluents such as starch, spray dried lactose, and microcrystalline cellulose tends to increase the hydrophilicity of the active ingredients and thus improve dissolution. Consequently, wet granulation

Figure 35-8. Effect of binders and granulating agents on dissolution rate of tablets.[5] **A.** Rate of dissolution of phenacetin from powder, granules, and tablets in diluted gastric juice (surface tension 42.7 dynes cm^{-1}, pH 1.85). ○, phenacetin powder; ▲, phenacetin granules; ●, phenacetin tablets. **B.** Dissolution rate of phenobarbital tablets in diluted gastric juice (surface tension 39.4 dynes cm^{-1}, pH 1.50). ● Gelatin binder, <open triangle> CMC, <open circle> Polyethylene glycol 6000.

Figure 35-9. Effect of starch content on dissolution rate.[4] ○, 5%; ●, 10%; ✕, 20% starch in granules.

Figure 35-10. Effect of lubricant on the dissolution rate of tablets.[6] **A.** Effect of magnesium stearate on dissolution rate of salicylic acid from rotating discs made from fine salicylic acid powder. ○, 3% magnesium stearate; ●, no lubricant added. **B.** Effect of lubricant on dissolution rate of salicylic acid contained in compressed tablets (formula A). ×, 3% magnesium stearate; ●, no lubricant; ○, 3% sodium lauryl sulfate.

was considered superior to a dry or double-compression procedure. Figure 35-11 shows the effect of different granulation methods on the dissolution rate of tablets.[15]

It must be noted that with the advent of newer tableting machines and materials, it becomes more evident that the critical formulation and proper mixing sequence and time of adding the several ingredients are the main criteria that affect the dissolution characteristics of the tablets, not the method of granulation.

EFFECT OF COMPRESSION FORCE ON DISSOLUTION RATE—In his early studies of the physics of tablet compression, T Higuchi (1953), pointed out the influence of compression force employed in the tableting process on the apparent density, porosity, hardness, disintegration time, and average primary particle size of compressed tablets. There is always a competing relationship between the enhancing effect due to the increase in surface area through the crushing effect and the inhibiting effect due to the increase in particle bonding that causes an increase in density and hardness and, consequently, a decrease in solvent penetrability. The high compres-

sion may also inhibit the wettability of the tablet due to the formation of a firmer and more effective sealing layer by the lubricant under the high pressure and temperature that usually accompanies a strong compressive force[13] (Fig 35-12). The curve profile of the compressive force of the tablet versus dissolution rate can take one of several shapes, as is observed in Figure 35-13.[8]

FACTORS RELATED TO THE DOSAGE FORM

DRUG EXCIPIENT INTERACTION—These interactions can occur during any unit operation, such as mixing, blending, drying, and/or granulating, resulting in a change in dissolution pattern of the dosage form in question.

The effect of magnesium stearate on the disintegration time of tablets containing either potato starch or sodium
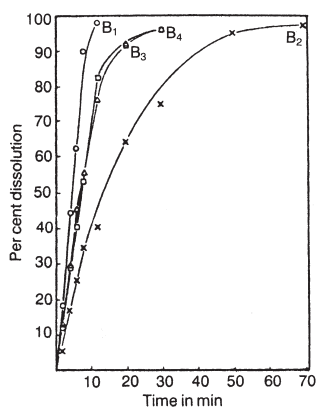
Figure 35-11. Effect of manufacturing process on the dissolution rate of tablets.[7] B_1, Direct compression with spray-dried lactose. B_2, Wet granulation with ethylcellulose and lactose. B_3, Acacia mucilage and lactose. B_4, Starch paste and lactose.
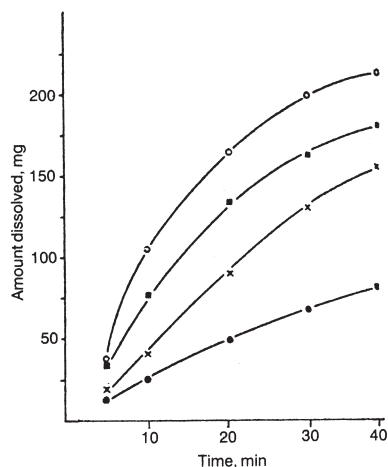
Figure 35-12. Effect of precompression pressure on the dissolution rate of salicylic acid contained in compressed tablets.[4] ●, 715 kg; ×, 1430 kg; □, 2860 kg; ○, 5730 kg pressure per cm.[2] (Average of five tablets each, formula D.)

starch glycolate was found to depend on the swelling characteristics of the disintegrants. These results were attributed to the formation of lubricant film during mixing, which resulted in an increase in disintegration time and thus delayed dissolution.[16]

It is essential that the formulator have a thorough understanding of these interactions so that the most appropriate excipients can be selected to enable the formulator to perform optimally. By minimizing, if not eliminating, these interactions, adverse effects on the performance of the final product can be avoided. It must also be noted that better process control is also possible with noninteracting drug-excipient interactions.

DEAGGREGATION—Deaggregation is often a prerequisite for dissolution. In such cases it can control dissolution. It was reported that two capsule formulations of sodium diphenylhydantoin showed significant deaggregation, dissolution, and thereby absorption rates. The formulation that deaggregated rapidly after the capsule shell was dissolved resulted in exposure of a larger surface area. This resulted in rapid dissolution at neutral pH but less rapid dissolution when both preparations were exposed to 0.1N hydrochloric acid. Aggregation of other formulation inhibited the conversion of most of its sodium salt to the free acid in acidic medium, whereas such conversion occurred readily with the rapidly disintegrating formulation. As a result, after neutralization of the medium, the latter dissolved and absorbed more readily and rapidly than did the former.
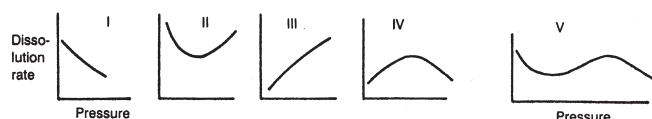
EFFECT OF TEST PARAMETERS ON THE DISSOLUTION RATE

ECCENTRICITY OF THE STIRRING DEVICE—USP 26/NF 21 specifies that the stirring shaft must rotate smoothly without significant wobble. Eccentricity can be measured with a machinist's indicator. It is measured in terms of total indicator reading (TIR), which determines the sum of the distance on both sides (180°C) of the axis of rotation.

GUIDING THE SHAFT—One must remember that the shaft of the stirring device extends about 6 in. beyond the chuck. An eccentricity of 0.005 in (0.11 mm) at a distance of 1 in (25 mm) from the chuck will be barely perceptible, but at 6 in it will amount to 0.30 in (0.75 mm), which is the maximum that can be tolerated.

The shaft mounting should not produce perfect concentricity but also allow for ease of vertical adjustment. That can best be obtained with a hollow drive shaft and chuck grip on the output with a guide on the other hand. They must be held to close concentricity tolerances with the axis of rotation of the drive tube. The further apart such guides, the better the probability of minimum wobble at the end of the shaft-provided the shaft is straight. The simple trigonometry is illustrated in the Figure 35-14.[1]

In the lower left view, the distance from the chuck to the basket is approximately 12 in (15 mm). If no guide bushing is used, any inherent eccentricity in the chuck is multiplied 12 times at the basket, which will certainly produce an eccentricity greatly exceeding the acceptable tolerances for eccentricity at the basket or paddle-if not when new, then after the chuck has been in use for a time.

In the lower right view, no guide bushing is used, but the shaft is supported at both the ends of the hollow shaft form A to A and the chuck is brought closer to the flask cover. The inherent eccentricity in the drive (A to A) can be held close at the factory, and the eccentricity of the basket cannot exceed it if the shaft is straight. In this case, the distance form A to A is about 6 in and matches the distance from the chuck head to the basket, also is about 6 in.
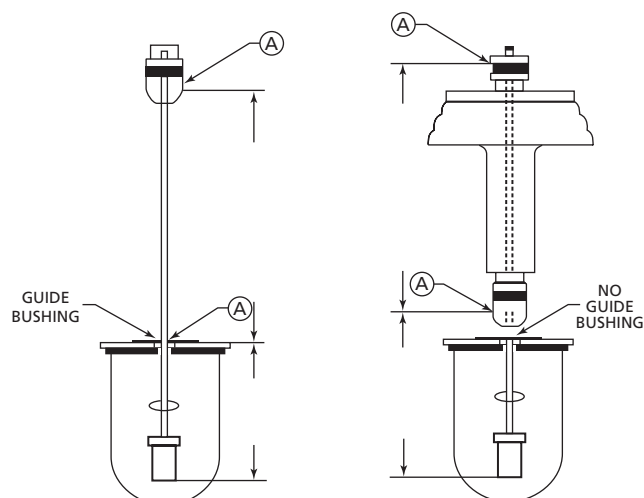
Even with a guide bushing, the system shown on the lower left is not recommended because the chuck may have a twist that is corrected by the bushing and that might cause a whip in the shaft with attendant vibration. These problems can be minimized by using a resilient grip in the chuck, such as a rubber "O" ring.

VIBRATION—Vibration is a common variable introduced into the dissolution system from myriad causes. It has the effect of changing the flow patterns of the liquid and of introducing unwanted energy to the dynamic system. Both effects may result in significant changes in the dissolution rates. The speeds of the rotational device selected by official compendium are 50 rpm or 100 rpm. Other speeds are specified for certain drugs. Precise speed control is best obtained with a synchronous motor that locks into the line frequency. Such motors are not only more rugged but are far more reliable. Periodic variations in rpm might result in possible disturbance in rotational devices, is commonly referred to as torsional vibration. Such vibration indicates a variation in the velocity of rotation for short periods of time-although the average velocity is well within ± 4% of the specified rate.

ALIGNMENT OF THE STIRRING ELEMENT—There are two important factors to be considered here. These are as follows:

Tilt—USP 26/NF 21 states that the axis of the stirring element shall not deviate more than 0.2 cm form the axis of the dissolution vessel, which defines centering of the stirring shaft to within ± 2 mm. It also constrains tilt. A series of tests suggest that tilt in excess of 1.5 (degrees) may increase dissolution rates using Method 2 from 2% to 25%,[17] which is still a significant variation. The user should be able to adjust his equipment to obtain alignment of the vertical spindles to within 1 (degrees) perpendicularly with the base of the drive to which the flasks are mounted. Such alignment cannot be ensured in the factory. Adjustments for perpendicularity must therefore be used in order to bring the equipment into alignment in its final position.[18]

Agitation Intensity—The degree of agitation, or the stirring conditions, is one of the most important variables to consider in dissolution. Given the background of various theories of dissolution, it is apparent that agitation conditions can markedly affect diffusion-controlled dissolution, because the thickness of the diffusion layer is

Figure 35-13. Different types of relations between compressional force of tablets and dissolution rate.

Figure 35-14. The rotating shaft should be supported at two places (A) to minimize wobble, shown by the two arrangements depicted.

inversely proportional to agitation speed. Wurster and Taylor[19] employed the empirical relationship

$$K = a(N)^b \tag{16}$$

where, N is the agitation rate, K the reaction (dissolution) rate, a and b are constants. For diffusion controlled processes, b = 1. Dissolution that is interfacial-reaction-rate-controlled will be independent of agitation intensity, and thus b = 0.

Agitation intensity within and between various *in vitro* dissolution testing devices can be varied by the dimensions and geometry of the dissolution vessel, volume of dissolution medium, and the degree of agitation or shaking. It is safe to predict that the two dosage forms having particles of differing sizes and densities will not experience identical dissolution system, even though the containers are being subjected to the same rate of rotation as of oscillation.

TEMPERATURE—Because drug solubility is temperature-dependent, careful temperature control during the dissolution process is very important and should be maintained within 0.5°. Generally, a temperature of 37° is always maintained during dissolution determinations. The effect of temperature variations of the dissolution medium depends mainly on the temperature/solubility curves of the drug and excipients in the formulation[20,21] (Fig 35-15).

For a dissolved molecule, the diffusion coefficient, D, depends on the temperature T according to the Stokes equation

$$D = kT/6\pi\eta r \tag{17}$$

where k is the Boltzmann constant and $6\pi\eta r$ is the Stokes force for a spherical molecule (η is the viscosity in cgs or poise units, and r is the radius of the molecule).
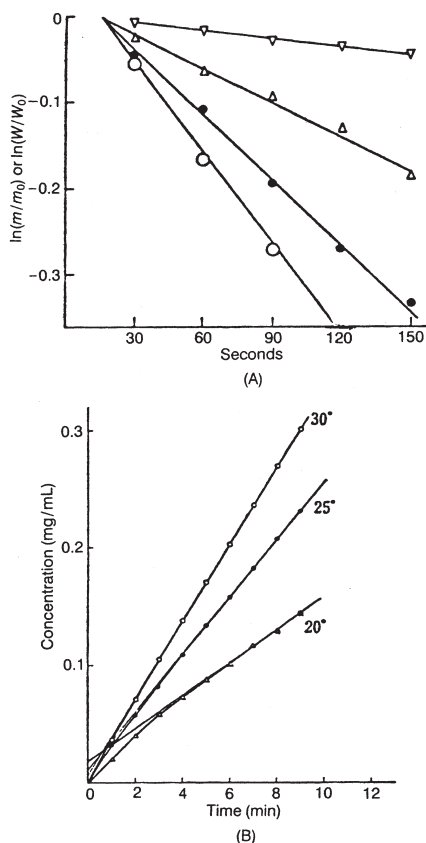
Figure 35-15. Effect of temperature on dissolution and disintegration rates of tablets.[15,16] **A.** Dissolution and disintegration curves according to Equations 1 and 2 for position II of the USP basket. ▽, dissolution at 10°; ○, dissolution at 20°; ●, dissolution at 30°; and ○, disintegration at 5°.[16] **B.** Dissolution of phenobarbital anhydrate at various temperatures (at 300 rpm).

DISSOLUTION MEDIUM

Selection of suitable fluid for dissolution testing depends largely on the solubility of the drug, as well as mere economics and practical reasons.

pH OF THE DISSOLUTION MEDIUM—Great emphasis and effort was first placed on simulating *in vivo* conditions, especially pH, surface tension, viscosity, and sink condition. Most of the early studies were conducted in 0.1N HCl or buffered solutions with a pH close to that of the gastric juice (pH ~ 1.2). The acidic solution tends to disintegrate the tablets slightly faster than water and thereby may enhance the dissolution rate by increasing the effective surface area. However, because of the corroding action of the acid on dissolution equipment, currently it is a general practice to use distilled water unless investigative studies show a specific need for the acidic solution to generate meaningful dissolution data. Another approach for avoiding the deleterious effects of hydrochloric acid is to replace it with acidic buffers, such as sodium acid phosphate, to maintain the required low pH.

SURFACE TENSION OF THE DISSOLUTION MEDIUM—Surface tension has been shown to have a significant effect on the dissolution rate of drugs and their release rate from solid dosage forms. Surfactants and wetting agents lower the contact angle and consequently improve penetration by the dissolution medium. Measurable enhancement in the dissolution rate of salicylic acid from an inert matrix was reported by Singh and co-workers when the contact angle, θ, was lowered from 92° (water) to 31° (using 0.01% dioctyl sodium sulfosuccinate)[22] (Fig 35-16). The surface tension also was correspondingly lowered from 60 to 31 dynes/cm. Similar findings were obtained in benzocaine studies when polysorbate 80 was used as the surface active agent[22] (see Fig 35-16).

Other studies conducted on conventional tablet formulations and capsules also showed significant enhancement in the dissolution rate of poorly soluble drugs when surfactants were added to the dissolution medium, even at a level below the critical micelle concentration, probably by reducing the interfacial
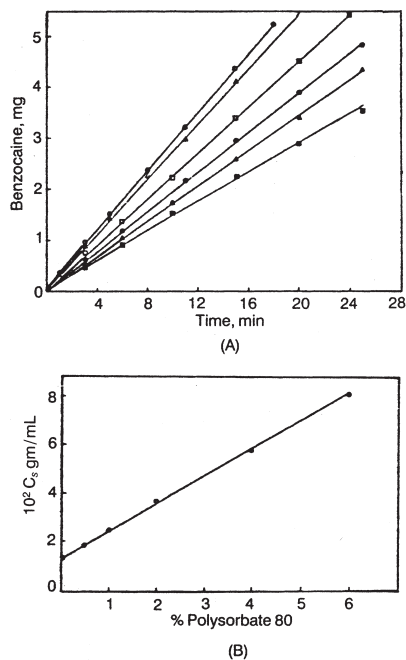
Figure 35-16. Effect of surfactants on dissolution rate.[17] **A.** Dissolution data for benzocaine in different concentrations of polysorbate 80 using the propeller-driven stirrer apparatus at a stirring speed of 150 rpm. Polysorbate conc: ○, 6%; △, 4%; □, 2%; ●, 1%; ▲, 0.5%; □, 0%. **B.** Solubilization data for benzocaine in different concentrations of polysorbate 80.

tension. Low levels of surfactants were recommended to be included in the dissolution medium as this seemed to give a better *in vivo* and *in vitro* correlation.

Finholt and Solvang compared the dissolution behavior of phenacetin and phenobarbital tablets in human gastric juice to that in dilute hydrochloric acid with and without various amounts of polysorbate 80 in the dissolution medium. The data showed that both pH and surface tension have significant influence on the dissolution kinetics of the drug studies. For example, they found that not only was the dissolution rate much faster in diluted gastric juice, but that it increased with decreasing particle size, whereas the opposite was the case when 0.1N HCl was used.

VISCOSITY OF THE DISSOLUTION MEDIUM—In case of diffusion-controlled dissolution processes, it would be expected that the dissolution rate decreases with an increase in viscosity. In the case of interfacial-controlled dissolution processes, however, viscosity should have little effect. The Stokes—Einstein equation describes diffusion coefficient, D, as a function of viscosity.

Braun and Parrott showed that the dissolution rate of benzoic acid is inversely proportional to the viscosity of the dissolution medium using various concentrations of sucrose and methylcellulose solutions[23] (Fig 35-17).

MISCELLANEOUS FACTORS

In addition to the factors discussed earlier, there are several other factors that can affect the dissolution characteristics of the drug product.

ADSORPTION—The adsorbent has an influence on the dissolution rate of a slightly soluble solid. It was also reported that the adsorbent is capable of increasing the dissolution rate observed in water under conditions of a decreased concentration gradient applying Nernst-Brunner film theory. Maximum dissolution rate can be obtained when a constant-concentration gradient is maintained. Adsorption isotherms can be employed

to calculate the approximate amount of adsorbent required to increase the slower dissolution rate.

SORPTION—The effect of water sorption on disintegration and dissolution properties, among other physical properties, of tablets containing microcrystalline cellulose was examined. It was concluded that water sorption from the atmosphere into the tablet containing microcrystalline cellulose is a very rapid first-order process, resulting in substantial changes in the physical properties. These changes are attributed to the breaking of the hydrogen bonds. The relative density of the tablets was found to decrease, resulting in increased disintegration time with increase in water sorption-rate constants. These changes were found to be irreversible.

HUMIDITY—In relation to the dissolution rate of a drug substance, humidity is usually associated with storage effects. Moisture has shown to influence the dissolution rate of many drugs from solid dosage forms. Environmental conditions to which dosage forms are exposed, moisture in particular, should be rigorously assessed if reproducible and reliable dissolution data are to be obtained. Additionally, humidity during the manufacture of the dosage forms should be carefully controlled to reproduce the quality of the product from batch to batch.

DETECTION ERRORS—Two most common variables leading to interlaboratory disagreement are the failure to use standards during analysis, and external vibration.[24] Extreme care must be exercised when laboratory methods are introduced into quality control to ensure that no part of the equipment interferes with sensitive determinations.

Despite the fundamental relationship between bioavailability and dissolution rate, the present evidence suggests that no single dissolution-rate test can be applied to all drugs. The possibility that a single test may be applied to drugs having similar physicochemical properties remains to be established. These observations are attributable, primarily, to the inability to assess and control the many variables affecting the dissolution process of a drug substance.

COMPENDIAL METHODS—When selecting apparatus for dissolution testing, routine quality control, new drug development, or complying with regulatory requirements, the analyst must follow the latest issue of compendia, including revisions. The modifications introduced in the dissolution testing methods during recent years are numerous that even revisions 2 or 3 years old may be outdated.

USP/NF Method 1 (Rotating Basket Method)—The USP/NF rotating basket method of dissolution testing essentially consists of a 1-in-diameter × 13/ 8-in-high stainless-steel 40-mesh wire basket rotated at a constant speed ranging between 25 and 150 rpm. It is immersed in 900 ml of dissolution medium in a vessel of 1000 ml capacity. The medium in the vessel is maintained at a constant temperature of 37 ± 0.5°C by means of a suitable water bath. The environment in which the apparatus is placed should not contribute significant motion, agitation, or vibration to the assembly. A fitted cover may be used to retard evaporation. The shaft is positioned so that its axis is not more than 2 mm at any point from the vertical axis of the vessel and rotates smoothly without any significant wobble (Fig 35-18).

The dosage unit is placed in a dry basket at the beginning of each test. Distance between inside bottom of the vessel and the basket is maintained at 25 ± 2 mm during the test.

In case of non-disintegrating dosage forms this apparatus is superior to Apparatus 2 since it constrains the dosage form in steady state fluid flow. This method may seem to be inferior for testing of dosage forms, which contain gums due to the clogging of screen matrix. In case of floating dosage forms this method performs well, but care should be taken that excipients do not clog the basket mesh.

USP/NF Method 2 (Rotating Paddle Method)—For all practical purposes the compendial specifications outlined for this method are identical to method 1 except that the paddle is substituted for the rotating basket.

The metallic or suitably inert, rigid blade and shaft comprise a single entity. The paddle and blade shaft may be coated with suitable inert coating. The dosage form is allowed to sink to the bottom of the vessel before rotation of the blade is started. This apparatus is frequently used for both disintegrating and non-disintegrating dosage form at 50 rpm. Other agitation speeds are acceptable with proper justification.
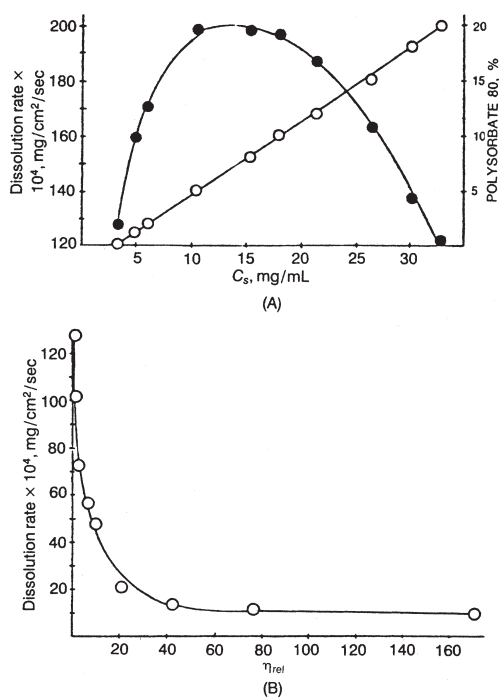
Figure 35-17. Effect of viscosity on dissolution rate.[18] **A.** Relationship of total solubility (C_s) of benzoic acid at 25° to dissolution rate and concentration of polysorbate 80. ●, rate; ○, concentration. **B.** Relationship of viscosity to dissolution rate of benzoic acid in aqueous methylcellulose solutions at 25°.

USP/NF permits variation in the paddle method involving the use of a helix of non-reactive material as a "sinker" for floating dosage forms. Anchoring accomplished by such a device has been severely studied (Fig 35-19).

USP/ NF Method 3 (Reciprocating Cylinder)—The assembly consists of a set of cylindrical, flat bottomed glass vessels; a set of glass reciprocating cylinders; stainless steel fittings (type 316 or equivalent) and screens that are made of suitable nonsorbing material and nonreactive material and that are designed to fit the top and bottoms of the reciprocating cylinders; and a motor and drive assembly to reciprocate the cylinders vertically inside the vessels and, if desired, index the reciprocating cylinders horizontally to a different row of vessels. The vessels are immersed in suitable water bath of any size that permits holding the temperature at 37 ± 0.5°C during the test. The components conform to the specifications as shown in the Figure 35-20 unless otherwise specified in the individual monograph.

One advantage of reciprocating cylinder is that gastrointestinal tract conditions can be easily simulated, as it is easy to make time dependent pH changes. This apparatus is most suitable for nondisintegrating (extended release) or delayed-release dosage (enteric coated) dosage forms.

USP Apparatus 4 (Flow-Through Cell)—The assembly consists of a reservoir and a pump for dissolution medium; a flow-through cell; a water bath that maintains dissolution medium at 37 ± 0.5°C. The pump forces the dissolution medium upwards through the flow-through cell. The pump has a delivery range between 240 and 960 ml/ hr, with the standard flow rates of 4, 8, and 16 ml/min. It must be volumetric to deliver constant flow independent of flow resistance in the filter device; the flow profile is sinusoidal with a pulsation of 120 ± 10 pulses per minute.

The components conform to the specifications as shown in the Figure 35-21 unless otherwise specified in the monograph.

The advantages of flow through cell apparatus most often cited are the ability to test drugs of very low aqueous solubility in the open loop mode and the ability to change the pH conveniently during the test. The disadvantage associated with it might be the operational difficulties of preparing large volumes of medium for operation in the open loop mode and the added time in the system set up and cleaning.

USP Apparatus 5 (Paddle Over Disk)—The Apparatus 2 is used, with the addition of a stainless steel disk assembly designed for holding the transdermal system at the bottom of the vessel. Temperature is maintained at 32 ± 0.5°C. A distance of 25 ± 2 mm between the paddle and blade and the surface of the disk assembly is maintained during the test. The vessel may be covered during the test to minimize evaporation. Disk assembly for holding the transdermal system is designed to minimize any 'dead' volume between the disk assembly and the bottom of the
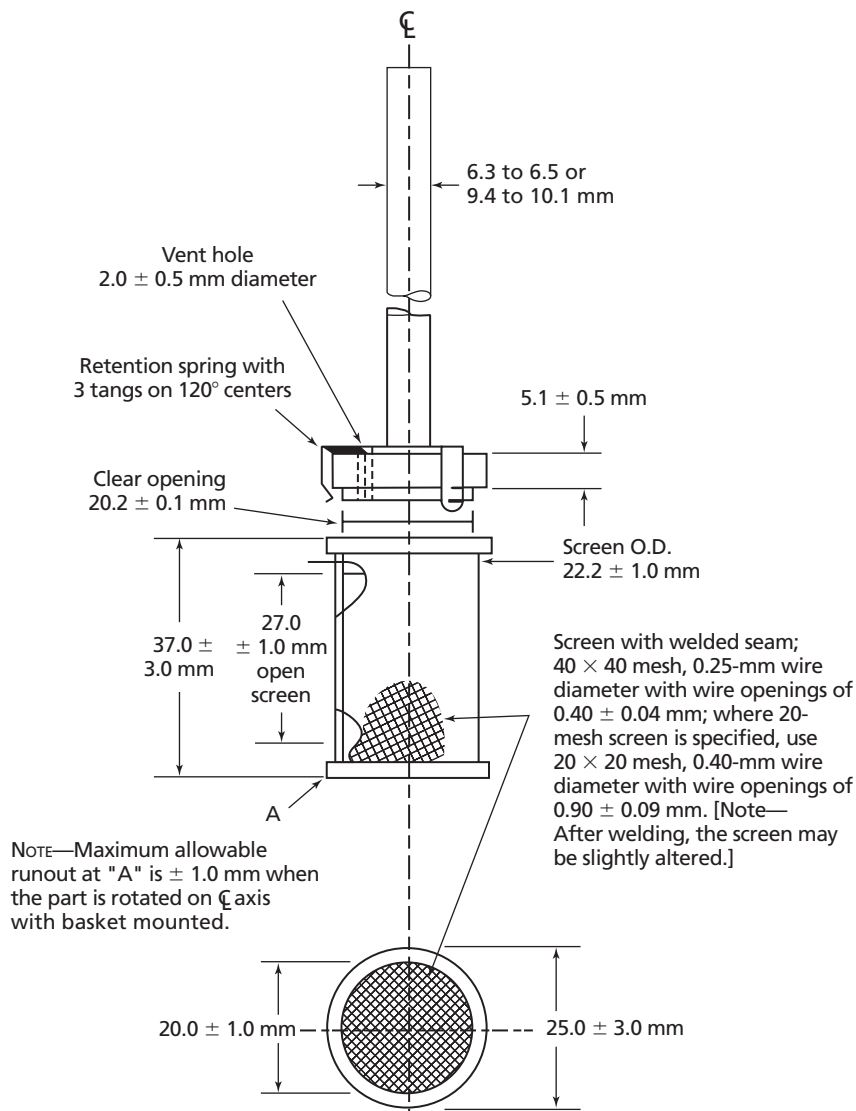

Figure 35-18. USP Apparatus 1.

9.4 to 10.1 mm diameter
before coating

NOTES —
(1) Shaft and blade material
    303 (or equivalent)
    stainless steel.
(2) A and B dimensions are
    not to vary more than
    0.5 mm when part is
    rotated on ₵ axis.
(3) Tolerances are ±1.0 mm,
    unless otherwise stated.

41.5 mm radius

1.2 mm radius

A

35.8 mm

B

19.0 mm
±0.5 mm

42.0 mm
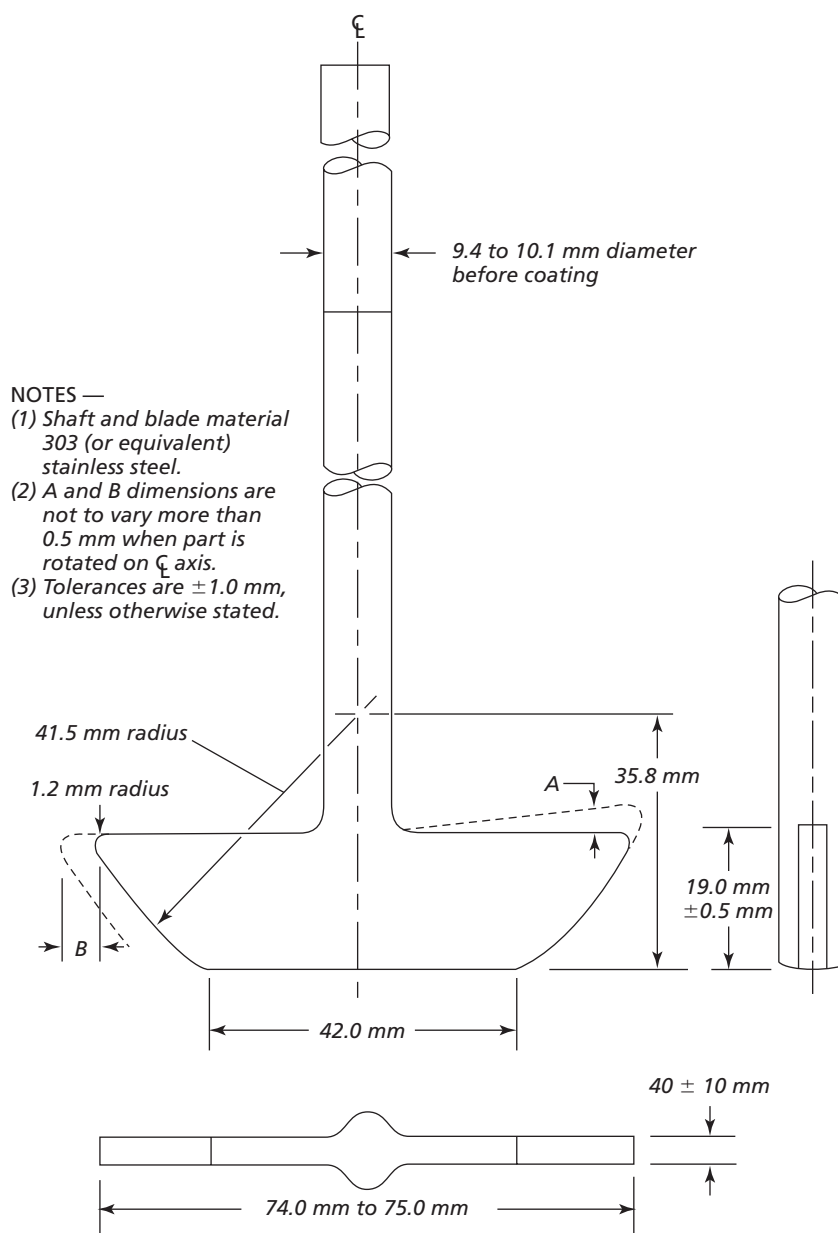
40 ± 10 mm

74.0 mm to 75.0 mm

Figure 35-19. USP Apparatus 2.

vessel. Disk assembly holds the system flat and is positioned such that the release surface is parallel with the bottom of the paddle blade. For more specifications refer to Figure 35-22.

USP Apparatus 6 (Cylinder)—The vessel assembly used is same as Apparatus 1, except the basket and the shaft is replaced with a stainless steel cylinder stirring element and to maintain the temperature at $32 \pm 0.5°C$ during the test. The shaft and cylinder components of the stirring element are fabricated of stainless steel to the specifications as shown in Figure 35-23. The dosage units are placed on the cylinder at the beginning of each test. The distance between the inside of the vessel and the cylinder is maintained at 25 ± 2 mm during the test.

USP Apparatus 7 (Reciprocating Cylinder)—The assembly consists of a set of volumetrically calibrated or tared solution containers made of glass or other suitable inert material, a motor and drive assembly to reciprocate the system vertically and to index the system horizontally to a different row of vessels automatically if desired, and a set of suitable sample holders. For details on specifications refer to the Figure 35-24.

DISSOLUTION OF IMMEDIATE RELEASE SOLID ORAL DOSAGE FORMS—*In vitro* dissolution tests for immediate release solid oral dosage forms, such as tablets and capsules, are used to (I) assess the lot-to-lot quality of a drug product; (II) guide development of new formulations; (III) ensure continuing product quality and performance.

For the drug approval process, it is essential to have the current knowledge about solubility, permeability, dissolution, and pharmacokinetics of a drug product. Based on drug solubility and permeability, the following Biopharmaceutical Classification System (BCS) is recommended in the literature[25]:

Case 1: High solubility-High permeability drugs
Case 2: Low solubility- High permeability drugs
Case 3: High solubility- Low permeability drugs
Case 4: Low solubility-Low permeability drugs

This classification can be used as a basis for setting *in vitro* dissolution specifications and *in vivo—in vitro* correlation
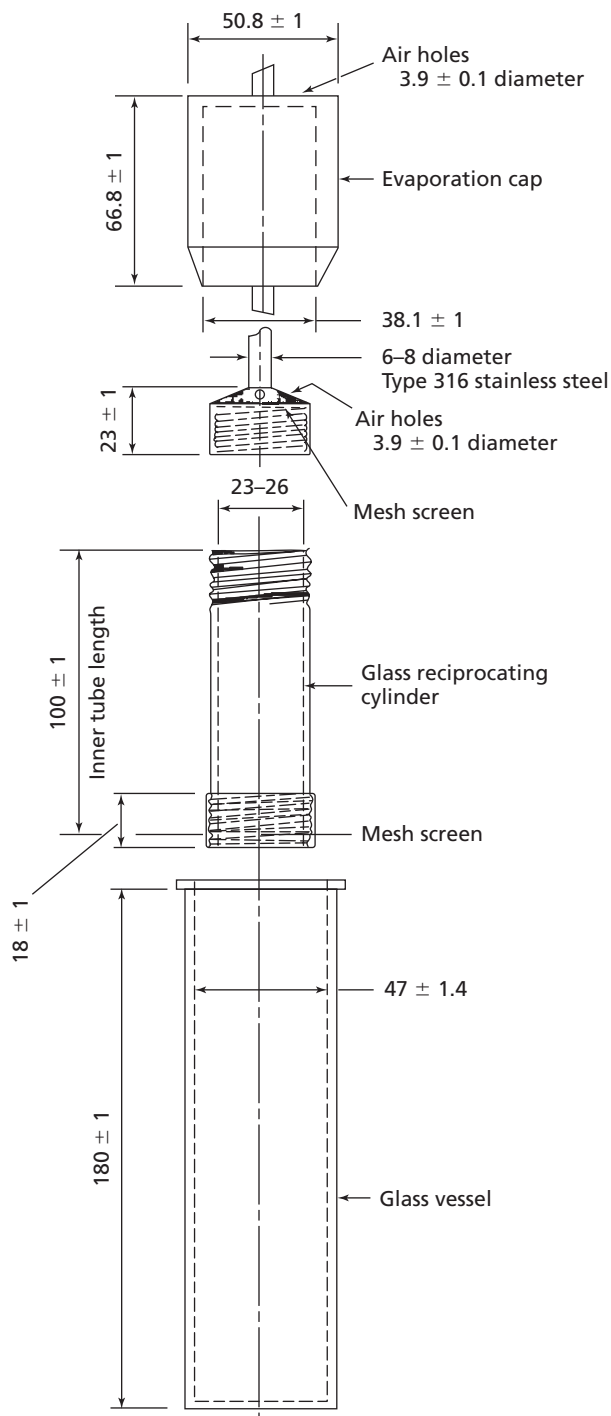
Figure 35-20. USP Apparatus 3. All measurements are expressed in mm unless noted otherwise.

(IVIVC). The BCS suggests that for high solubility, high permeability (Case 1) drugs and in some cases for high solubility, low permeability (Case 3) drugs, 85% dissolution in 0.1N HCl in 15 minutes can ensure that the bioavailability is not limited by dissolution. In case of low solubility, high permeability drugs (Case 2), drug dissolution may be the rate-limiting step for drug absorption and an IVIVC may be expected. A dissolution profile in multiple media is recommended for drug products in this category. In case of high solubility, low permeability drugs (Case 3), permeability is the rate controlling step and a limited IVIVC may be possible, depending on the relative rates of dissolution and intestinal transit. Drugs in low solubility, low permeability (Case 4) present significant problems for oral drug delivery.

DISSOLUTION OF ORALLY DISINTEGRATING TABLETS [26]—Orally disintegrating tablets (ODT) are solid dosage forms that disintegrate in the oral cavity leaving an easy to swallow residue. ODT in general have high porosity, low density, and low hardness. The time for disintegration for ODT is usually considered to be less than 1 minute. Development of dissolution methods for ODT is comparable to the approach taken for conventional tablets except when the tablets utilize taste masking. Media that can be used are 0.1 N hydrochloric acid, and pH 4.5 and 6.8 buffers. The most commonly used apparatus for running dissolution test for ODT is USP Apparatus 2 (Paddle method) with a paddle speed of 50 rpm. USP Apparatus 1 is less frequently used due to the physical properties of these tablets, as the tablet fragments or disintegrated tablet masses may become trapped in the basket yielding poorly reproducible dissolution profiles. Since dissolution for ODT is very fast, slower speeds are employed. In case of tablets exceeding 1 gram and containing relatively denser particles larger mounds may be produced on dissolution, which may be prevented by using higher paddle speeds. These two situations expand the suitable range to 25–75 rpm.

DISSOLUTION OF TOPICAL DOSAGE FORMS [27]—Drug-release studies from gels, creams, and ointments are becoming an important step both during the developmental stages of new formulations and as a routine quality control test for assuring the uniformity of the finished product. Also these studies often can provide useful information on some physico-chemical parameters involved in the *in vivo* percutaneous absorption, such as the diffusion coefficient and the solubility of the drug in the specific vehicle used.

Although many investigators have conducted drug release-rate studies from topical dosage forms, it appears that no single apparatus or procedure has yet emerged as the most favored, or to be accepted widely as a quasi-standard for others in the field. According to FDA guidelines the most commonly used method is as follows:

In vitro dissolution method for topical dosage forms is based on an open chamber diffusion cell system such as a Franz cell system, fitted usually with a synthetic membrane. The test product is placed on the upper side of the membrane in the open donor chamber of the diffusion cell and a sampling fluid is placed on the other side of the membrane in a receptor cell. Diffusion of drug from the topical product to and across the membrane is monitored by assay of sequentially collected samples of the receptor fluid.

Aliquots removed from the receptor phase can be analyzed for drug content by high-pressure liquid chromatography (HPLC) or other analytical methodology.

DISSOLUTION OF SUSPENSIONS—Although most dissolution studies during the last two decades have concentrated on tablets and capsules, some studies have pointed to the importance of the dissolution characteristics of drugs administered in suspension. This hardly is surprising, as suspensions are similar to the disintegrated form of tablets and capsules; if dissolution has become a priority for these formulations; it is logical to extend its concept to suspensions. Indeed, several studies have shown that the absorption of several poorly soluble drugs administered in suspension formulations is dissolution rate—limited.

Such *in vivo/in vitro* correlation studies have confirmed the importance and the viability of dissolution rate determinations of suspensions as a discriminative test for rapid screening of new formulations and to control lot-to-lot variability within the same manufacturer and between different commercial manufacturers. In general, most of the dissolution apparatuses that have been described for tablets and capsules easily could be used for suspensions.

The USP Apparatus 2 (Paddle) has been used frequently at a rotation speed between 25 to 50 rpm. However, the rotating filter apparatus by Shah has gained wide acceptance for sus-
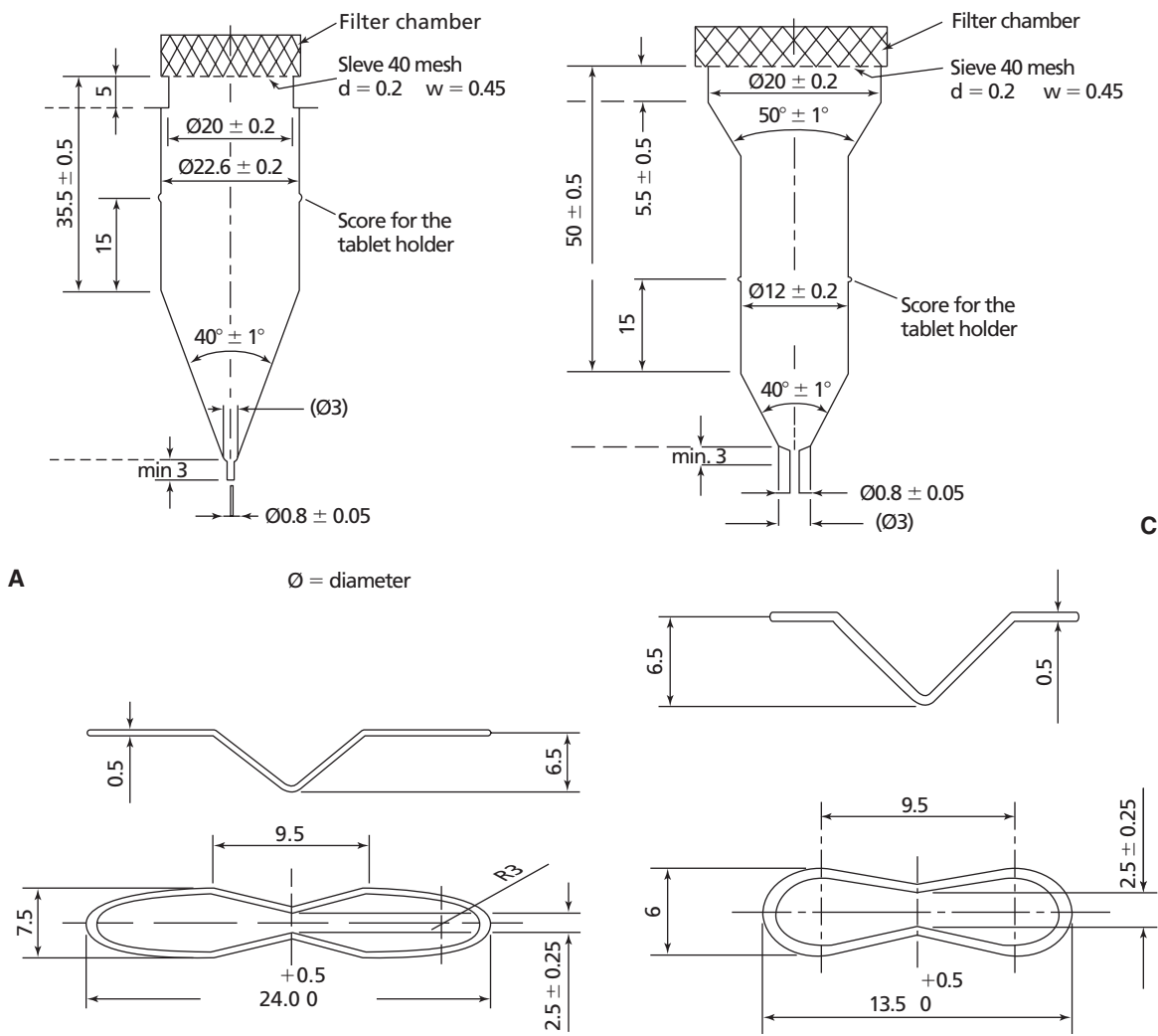
Figure 35-21. USP Apparatus 4. **A.** Large cell for tablets and capsules. All measurements are expressed in mm unless noted otherwise.**B.** Tablet holder for the large cell. All measurements are expressed in mm unless noted otherwise. **C.** Small cell for tablets and capsules. All measurements are expressed in mm unless noted otherwise. **D.** Tablet holder for small cell. All measurements are expressed in mm unless noted otherwise.
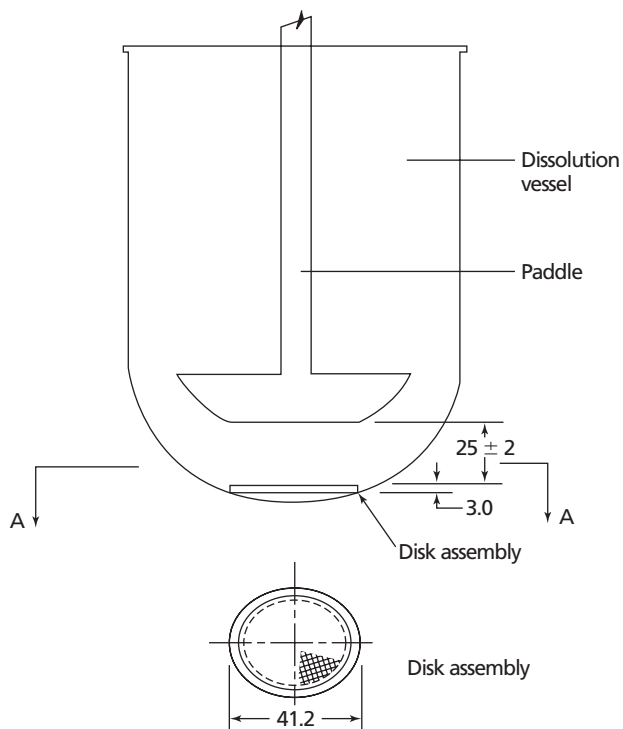
Figure 35-22. USP Apparatus 5 (paddle over disk). All measurements are expressed in mm unless noted otherwise.
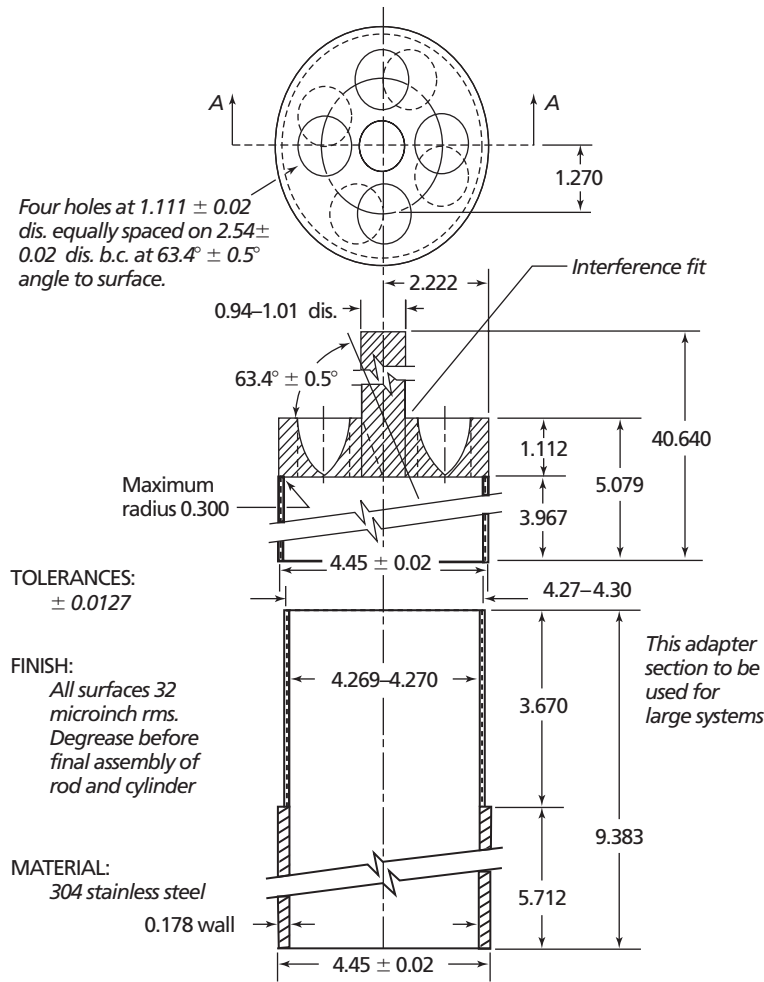
Figure 35-23. USP Apparatus 6. All measurements are expressed in cm unless otherwise noted.

pensions because it provides mild laminar liquid agitation, and it also functions as an *in situ* nonclogging filter. Sufficient volume of the dissolution medium should be used to maintain sink condition (about 900–1000 mL), and a temperature of 37° should be maintained.

DISSOLUTION OF SUPPOSITORIES—Although most of the early work on suppositories has been concerned with their physical characteristics, such as softening and liquefaction ranges, homogeneity, smoothness, and neutrality, several reports appeared in the early literature pointing to the direct correlation between their efficacy and the release characteristics of the active ingredients. It has been reported that fatty bases, such as the popular cocoa butter, tend to release hy-

Figure 35-24. USP Apparatus 7. Transdermal system holder-angled disk

drophobic drugs, which are highly soluble in the oily base, very slowly. Emulsification of the fatty base significantly improved the drug-release rate. Incorporation of surface-active agents was found to improve the release rate of water-soluble drugs from the fatty suppository base dramatically.

Although many investigators have conducted extensive research on the release of drugs from suppositories, no single method or apparatus design has yet emerged as the standard procedure for the pharmaceutical laboratory. Many methods for the determination of the dissolution rate of suppositories are based on the dialysis technique, where the suppository is placed in a dialyzing bag made of special membrane or cellophane material. The bag is placed in a beaker or wide-mouth bottle containing a known volume of distilled water, and the concentration of the drug outside of the bag is measured as a function of time.

A slight variation of the basket method of the USP Dissolution Apparatus 1 also is used frequently. Hanson Research markets a basket apparatus for suppository dissolution testing. Hanson's modified basket uses slots instead of mesh to provide a suitable porosity. The use of such a basket avoids the blocking of the mesh opening of the regular USP basket when oil-based suppositories are used. The system also has the advantage of being capable of testing suppositories that float or have such low specific gravity that it interferes with the flow dynamics in the paddle method.

DISINTEGRATION AND DISSOLUTION OF NUTRITIONAL SUPPLEMENTS[28]**—**USP 26/NF 21 contains a section on Nutritional Supplements. One of the specifications ap-
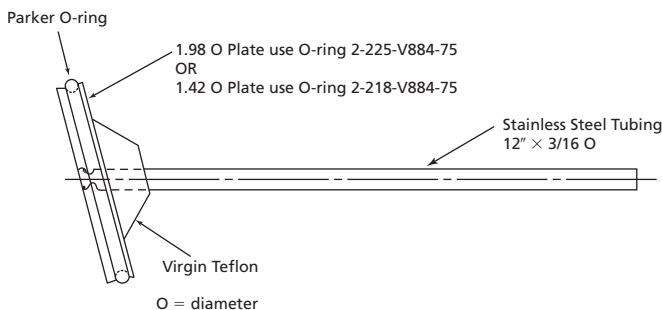
pearing in the monographs for some of the supplement dosage forms is Disintegration and Dissolution, 2040. The dissolution procedures for the nutritional supplements use Apparatus 1 and Apparatus 2 and require measurement of one vitamin and folic acid (if applicable) and one mineral (if applicable). Oil-soluble vitamins are exempted from the dissolution requirement.

DISSOLUTION OF MODIFIED RELEASE DOSAGE FORMS[27]—

Extended Release—In addition to application/compendial release requirements, multipoint dissolution profiles should be obtained in three other media, for example, in water, 0.1N HCl, and USP buffer media at pH 4.5, and 6.8 for the drug product. Adequate sampling should be performed, for example, at 1, 2, and 4 hours and every 2 hours thereafter until either 80% of the drug from the drug product is released or an asymptote is reached. A surfactant may be used with appropriate justification.

Delayed Release—In addition to application/compendial release requirements, dissolution tests should be performed in 0.1 N HCl for 2 hours (acid stage) followed by testing in USP buffer media, in the range of pH 4.5-7.5 (buffer stage) under standard (application/compendial) test conditions and two additional agitation speeds using the application/compendial test apparatus (three additional test conditions). If the application/compendial test apparatus is the rotating basket method (Apparatus 1), a rotation speed of 50, 100, and 150 rpm may be used, and if the application/compendial test apparatus is the rotating paddle method (Apparatus 2), a rotation speed of 50, 75, and 100 rpm may be used. Multipoint dissolution profiles should be obtained during the buffer stage of testing. Adequate sampling should be performed, for example, at 15, 30, 45, 60, and 120 minutes (following the time from which the dosage form is placed in the buffer) until either 80% of the drug from the drug product is released or an asymptote is reached.

DISSOLUTION PROFILE COMPARISONS

In the presence of minor changes, single point dissolution tests have been employed in evaluating scale-up and post approval changes.[29] For major changes, a dissolution profile comparison performed under identical conditions for the product before and after the change is recommended. Dissolution profile comparison may be carried out using the model dependent or model independent methods. One such model independent approach has been explained in the following paragraph.

Model Independent Approach Using A Similarity Factor

This approach uses a difference factor (f_1) and a similarity factor (f_2) to compare the dissolution profiles. The difference factor (f_1) calculates the percent (%) difference the two curves at each time point and is a measurement of the relative error between the two curves:

$$f_1 = \left\{ \left[\sum_{t=1}^{n} |R_t - T_1| \right] / \left[\sum_{t=1}^{n} R_t \right] \right\} \cdot 100$$

Where n is the number of time points, R_t is the dissolution value of the reference batch (prechange) at time t, and T_t is the dissolution value of the test (postchange) batch at time t.

The similarity factor (f_2) is a logarithmic reciprocal square root transformation of the sum of squared error and is a measurement of the similarity in the percent (%) dissolution between the curves.

$$f_2 = 50 \cdot \log \left\{ \left[1 + \left(\frac{1}{n} \right) \sum_{t=1}^{n} |R_t - T_t|^2 \right]^{-0.5} \cdot 100 \right\}$$

In order to calculate the difference and similarity factor, first the dissolution profile should be done for 12 units each of the prechange and the postchange products. The difference factor (f_1) and similarity (f_2) can be calculated using the mean dissolution values from both curves at each time interval. For the curves to be considered similar, f_1 values should be close to 0,

and f_2 values close to 100. This model independent method is most suitable for dissolution profile comparison when three or four more dissolution time points are available.

Automation in Dissolution Testing

Due to the large amount of testing required in determining dissolution rate of drugs, automation of the process seemed almost a necessity and not simply a convenience to the analyst. Also, because of modular nature of the dissolution apparatus, automation can be accomplished easily in different ways and by various techniques.

At present, however, the setup of the apparatus, media preparation, and introduction of the dosage forms mostly are done manually. The rest of the process—including the withdrawal of samples, maintenance of a certain pH or of sink conditions, assay performance, and data acquisition and calculations—is in most cases fully automated. The automation process not only saves money, time, and effort on the part of the analyst, but more significantly it improves the overall reliability and enhances the reproducibility of testing procedures.

Several commercial companies have also introduced semi- and fully automated dissolution systems. Some of these are the Hanson Research Dissolution System (Northridge, CA; Dissoette and Dissograph apparatuses), Technicon (Tarrytown, NY; Sasdra apparatus), and Applied Analytical (Wilmington, NC).

Millipore's Waters Chromatography Division has introduced a fully automated dissolution system using a Waters pump, detector, and autosampler combined with a Hanson Research's dissolution bath and sample transfer system. Samples are analyzed by HPLC, which provides better specificity than ultraviolet (UV) methods of analysis.

Hewlett-Packard manufactures a fully automated dissolution-sampling and UV analysis system that can analyze samples from three dissolution baths. One such system is model 2100 C (Fig 35-25) dissolution test system.[30] The Model 2100C combines enhanced features with advanced communications to ensure reproducibility and control throughout the dissolution testing process. The operations are technician-friendly. Its convenient vessel layout simplifies manual or automatic sampling. It has built-in height adjustment and permanent centering which reduces operator errors. It also has chuckless spindles to reduce setup time. The precision control of variables provides most accurate and repeatable results. The unique water bath flow characteristics maintains vessel temperature to better than $\pm$ 0.1°C. It has versatile SystemLink™ for PC communication and printer output, and can be configured for use both with basket and paddles.
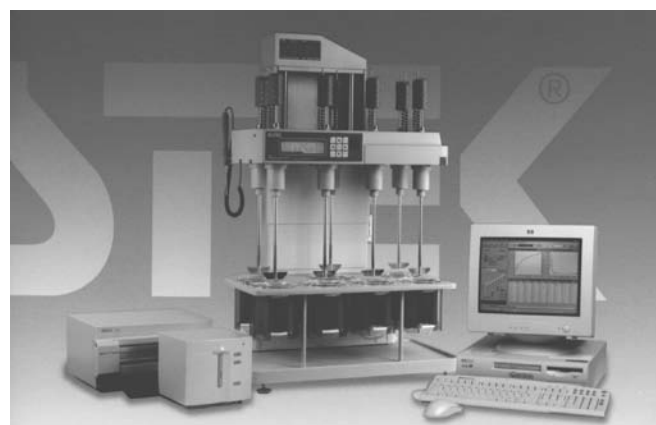

Figure 35-25. Distek dissolution system.

Validation of Dissolution Method

In general, the approach to validation of a dissolution method is similar to that of any other method. The following discussion briefly summarizes the approach to dissolution assay validation.[31]

LINEARITY, FILTER BIAS, AND RECOVERY STUDIES—The linearity of the detection method, the filter bias, and the recovery of drug from dissolution fluid containing placebo should be determined. System suitability tests for UV-Vis and chromatographic methods are also identified at this stage.

PRECISION AND RUGGEDNESS—Precision testing of the dissolution method should be performed on at least two lots of six tablets each on 2 days. The average of each run, as well as the standard and relative standard deviations, should be computed. Precision of the dissolution method is usually expressed as the standard deviation for a data set obtained on a single day.

EFFECT OF DISSOLVED GASES—Air dissolved in the media may form bubbles that in turn could coat the tablets or other dosage form. This is the most likely to happen as the medium is heated to test temperature (37°C). The coating can affect the drug release by altering the dissolution and disintegration or dissolution of the tablet. Accordingly, the effect of deaeration on the dissolution rate should be evaluated or deaerated medium should be specified in the procedure. Effective methods of deaeration include vacuum filtration, helium sparging, hot water placed under a vacuum with or without sonication, and the use of a commercially available medium dispensing device.[32]

AUTOMATION—Validation for automated systems is the same as for manual sampling. A simple experiment should be done in order to verify that the drug does not adsorb to the apparatus tubing and to quantify system carryover.

REFERENCES

1. Banakar UV. *Pharmaceutical Dissolution Testing*, 1st ed. New York: Marcel Dekker, 1991.
2. Carstensen TJ. *Dissolution—State of the Art 1982*. (Proc 2nd WI Update Conf) Madison, WI: Extension Services in Pharmacy, University of Wisconsin, 1982.
3. Morrison AB, Campbell JA. *J Pharm Sci* 1965; 54(1):1–8.
4. Wagner JG. *Biopharmaceutics and Relevant Pharmacokinetics*, 2nd ed. Hamilton: Drug Intelligence Publications, 1971, pp 190–196.
5. Abdou H. *Dissolution, Bioavailability and Bioequivalence.* Easton, PA: Mack, 1989, Chap 2.
6. Cox D, Douglas C, Furman W, et al. *Pharm Tech* 1978; 2(4):40–53.
7. Viegas TX, Curatella LVW, Brinker G. *Pharm Tech* 2001; 44–53.
8. Finholt P. In: Leeson LJ, Carstensen TJ, eds. *Dissolution Technology*. Washington, DC: APhA, 1974, p 108.
9. Aguiar AJ, Krc J, Kinkel AW, et al. *J Pharm Sci* 1967; 56(7):847–853.
10. Solvang S, Finholt P. *J Pharm Sci* 1970; 59(1): 49–52.
11. Jaminet F, Delattre L, Delporte JP. *Pharma Acta Helvetiae* 1969; 44(7):418–432.
12. Alam AS, Parrott EL. *J Pharm Sci* 1971; 60(2):263–266.
13. Levy G, et al. *J Pharm Sci* 1963; 52:1047.
14. Levy G, Gumtow RH. *J Pharm Sci* 1963; 52:1139.
15. Marlowe E, Shangraw R. *J Pharm Sci* 1967; 56:498.
16. Murthy KS, Samyn JC. *J Pharm Sci* 1977; 66(9):1215–1219.
17. Hanson W, Hanson R. *Pharm Tech* 1979: 3(3):42–50.
18. Thakker K, Naik N, Gray V, et al. *Pharm Forum* 1980; 6:177–185.
19. Wurster DE, Taylor PW. *J Pharm Sci* 1965; 54(5):670–676.
20. Nogami H. *Chem Pharm Bull* 1969; 17:499.
21. Carstensen TJ, et al. *J Pharm Sci* 1980; 69:291.
22. Singh P, et al. *J Pharm Sci* 1968; 57:959.
23. Braun R, Parrott E. *J Pharm Sci* 1972; 61:175.
24. Cartwright AC. *J Pharm Pharmacol* 1979; 31:434–440.
25. Amidon GL, Lennernas H, Shah VP, et al. *Pharm Res* 1995; 12:413–420.
26. Klancke J. *Dissolution Technologies* 2003; 10(2):6–8.
27. http://www.fda.gov/cder/guidance/1447fnl.pdf
28. US Pharmacopeia 26/ National Formulary 21
29. Moore JW, Flanner HH. *Pharm Tech* 1996; 20(6):64–74.
30. www.distek.com
31. Skug JW, et al. *Pharm Tech* 58, 1996.
32. Rohrs BR, Stelzer D J. *Dissolution Technologies* 1995; 2(2).

Pharmaceutical Manufacturing

Linda Felton, PhD, BSPharm, RPh
Associate Professor of Pharmaceutics
University of New Mexico
College of Pharmacy
Albuquerque, NM

Separation

Garnet E Peck, PhD

Separation may be defined as an operation that brings about isolation and/or purification of a single chemical constituent or a group of chemically related substances. Most medicinal agents require some degree of purification before being incorporated into desirable dosage forms. Many times the analysis of pharmaceutical preparations requires separation of the chief constituent from other formulation constituents before quantitative measurement can be made.

Although the problems of separation are the concern chiefly of pharmaceutical manufacturers, at times they may be encountered also by the pharmacist in the prescription laboratory; hence, all pharmacy practitioners should have knowledge of the underlying principles and the techniques employed in the basic processes of separation.

The processes of separation may be divided into two general categories—simple and complex—depending on the complexity of the method used.

Simple processes bring about separation of constituents through a single mechanical manipulation. Some examples of this type are the use of

- A separatory funnel or pipette to separate two immiscible liquids such as water and ether
- A distillation process to separate two miscible liquids such as benzene and chloroform
- A garbling process to separate solids
- Centrifugation, filtration, and expression processes to separate solids from liquids

Processes in this category are limited usually to separations of relatively simple mixtures or solutions.

Complex processes usually require formation of a second phase by addition of either a solid, liquid, or gas plus mechanical manipulation to bring about effective separation. One example is the separation of aspirin (acetylsalicylic acid) from salicylic acid. In this mixture, salicylic acid is considered to be an impurity, and to separate the impurity from the desired constituent, a suitable solvent is added to the mixture for the purpose of recrystallizing only the acetylsalicylic acid. The contaminant remains in solution and is removed in the filtrate during the filtration process.

Only selected processes involving separations will be covered in this chapter. Other methods are discussed in such chapters as *Complex Formation* (Chapter 14), *Colloidal Dispersions* (Chapter 21), *Coarse Dispersions* (Chapter 22), and *Chromatography* (Chapter 33).

COUNTERCURRENT DISTRIBUTION

Countercurrent Distribution (CCD) may be defined as a series of liquid-liquid extractions (immiscible solvents) conducted in a multiple-tube apparatus in which one phase is permitted to advance to the next tube in the series independently of the other phase.[1] The separation of the components in the mixture depends on the distribution coefficients of each of the components, the volume of the solvents used, and the number of transfers taken.

Some important applications of CCD in the pharmaceutical sciences are

- The isolation and purification of chemicals and biochemicals that might otherwise be damaged by the extremes of temperature or pH that occur during the separation processes
- The separation of a crude plant extract into its various chemically related fractions as a preparative step
- The determination of purity and homogeneity of chemicals and medicinal agents
- The characterization of substances extracted from biochemical systems in studies determining the metabolic or biologic disposition of drugs

Separation using CCD is based on Nernst Law. According to this law, when two practically immiscible solvents are in contact with each other and a substance that is soluble in each is added, the substance distributes itself in such a way that at equilibrium and at a given temperature the ratio of the concentrations of the two solutions is a constant. Strictly speaking, it is the activity ratio rather than the concentration ratio that remains constant. For most purposes, however, concentration values give satisfactory approximations.

When the ratio of concentrations expresses a distribution value for a single chemical species, the constant is designated as a partition coefficient or distribution coefficient, K, and may be expressed mathematically as

$$K = C_u/C_l \tag{1}$$

In this expression C_u and C_l represent concentrations in the upper and lower phases, respectively. There is no accepted convention to date, and the distribution coefficient could just as well be expressed as the reciprocal of Equation 1, C_l/C_u.

In actual practice one deals with and measures total analytical concentrations; thus, more than one chemical species usually is present in each phase. This type of distribution between solvents is called the partition ratio and is defined mathematically as $K_p = C_u/C_l$, where C_u and C_l represent total analytical concentrations of the chemical in the upper and lower phases, respectively. An example would be the distribution of benzoic acid between benzene and water. In the aqueous phase, benzoic acid would be present both in the ionized (A^-) and un-ionized form (HA). In benzene, benzoic acid would be present in the un-ionized form (HA) and in the dimerized form $(HA)_2$. The ratio expressing total benzoic acid in the organic phase and total benzoic acid in the aqueous phase is the partition ratio or the apparent distribution coefficient, K_p.

Although the purpose of using CCD is to bring about the separation of two or more substances, the basic principles of operation are best introduced by first considering the distribution pattern of a single solute in the two immiscible solvents.

1. Assume that the solute under consideration has a distribution coefficient of unity when distributed between chloroform and buffer solution and that there are no deviations from Nernst's law of distribution due to molecular association, dissociation, ionization, or chemical reactions.

2. Consider six containers such as 250-mL glass-stoppered Erlenmeyer flasks, each holding 50 mL of chloroform (lower phase) as shown in Figure 36-1 (Row A). Add to container No 0, 100 mg of solute under consideration dissolved in 50 mL of buffer solution, and shake until equilibrium has been established. Because equal volumes of solvent are used and the distribution coefficient of solute in these two solvents is unity, the solute at equilibrium will distribute itself in such a way that one-half is found in each of the upper and lower phases (Row B). Because 100 mg was originally present, 50 mg will be found in both layers of Container 0 (Row B).

3. Transfer the upper phase of Container 0 holding 50 mg of solute to Container 1 (Row B) and add fresh buffer solution to Container 0 (Row B). Shake both containers until equilibrium has been established. At equilibrium the quantity of solute in each phase of Containers 0 and 1 (Row C) will be 25 mg.

4. Transfer the upper phase of Container 1 (Row C) to Container 2 (Row C), and the upper phase of Container 0 (Row C) to Container 1. Add fresh buffer solution to Container 0 (Row C) and shake all three containers until equilibrium has been established. At equilibrium the quantity of solute (25 mg) in Container 2 (Row D) will have distributed itself so that one-half (12.5 mg) is in the upper phase and one-half (12.5 mg) is in the lower phase. Because 25 mg of solute was transferred to Container 1 from Container 0, 25 mg of solute will be present in each phase of Container 1 (Row D). The quantity (25 mg) of solute in Container 0 will distribute itself between the chloroform layer and freshly added buffer solution so that one-half (12.5 mg) will be present in each layer (Row D).

Continue this general procedure of transferring the upper phases of Containers 0, 1, and 2 to Containers 1, 2, and 3, respectively; then add fresh buffer to Container 0. Shake the four flasks until equilibrium is established. A distribution is obtained as shown in Row E. Continuing in a like manner will give a distribution as shown in Row F.
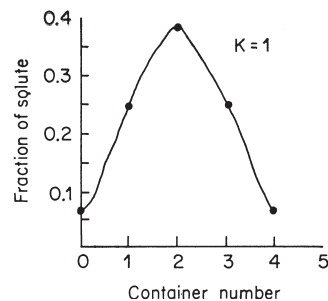

Figure 36-2. Distribution of solute after four transfers.

A plot of the fraction of solute in each container versus container number is shown in Figure 36-2. The significance of this curve is that the distribution of the solute shows a peak in which the maximum is located in a specific container and the location of the peak container is a function of the partition coefficient. Hence, it can be seen that two or more solutes with different K values can be separated effectively after the passage of a mixture through many tubes (usually 25 or more, depending upon K values) in a CCD apparatus.

Figure 36-2 illustrates the distribution of a solute after only four transfers. In actual practice between 8 and 2000 containers or tubes usually are used in multiple extractions of this kind. The tubes are connected in series in a train and are rocked simultaneously rather than individually to bring about distribution of solutes between the two phases. The device also permits the transfer of upper phases to the next tube in series, in one operation. A device of this type is called a countercurrent distribution apparatus.

To study the fraction of a given solute present in each tube r, after n number of transfers, it is convenient to use Equation 2,

$$f_{n,r} = \frac{n!}{r!(n-r)!}\left(\frac{1}{1+KR}\right)^n (KR)^r \qquad (2)$$

where K is defined as the partition coefficient and R is defined as the ratio of the volume of the upper phase to the volume of the lower phase, (V_u/V_l).

The use of Equation 2 is illustrated as follows: Calculate the fraction of solute in tubes no 0, 1, 2, 3, and 4 after four transfers are made in a CCD apparatus using equal volumes of upper and lower phases. The K value for the solute in the solvent system is assumed to be 1.0 in this example.

For Tube 3,

$$f_{4,3} = \frac{4!}{3!(4-3)!}\left(\frac{1}{1+1}\right)^4 (1)^3 = 0.25$$

By similar calculations the fraction of solutes in Tube 0, 1, 2, and 4 is found to equal

$$f_{4,0} = 0.0625; \; f_{4,1} = 0.25; \; f_{4,2} = 0.375; \; f_{4,4} = 0.0625$$

The distribution of solute using Equation 2 is shown in Figure 36-2.

When a large number of transfers (50) are made and K is near unity it is more convenient to use a Gaussian treatment[2] to calculate the fraction of solute in a particular tube. The appropriate equations are

$$y_x = \frac{1.00}{\sqrt{2\pi nKR/(KR+1)^2}}\exp\left\{-\left(\frac{x^2}{2nKR/(KR+1)^2}\right)\right\} \quad (3)$$

$$\tau_{max} = \frac{nKR}{KR+1} \qquad (4)$$

where y_x represents the fraction of solute with distribution coefficient K in the tube that is x distant from the peak tube; exp is the exponent of the base e, ex, exp2 = e^2; π = 3.14; K, R, and n are terms that have been defined previously and r_{max} represents the number of the tube containing the maximum amount of solute.

Container no, r	0	1	2	3	4	5	
CHCl$_3$ only in each container							A
Initial Distribution (n=0)	50 / 50						B
Distribution after 1st transfer (n=1)	25 / 25	25 / 25					C
Distribution after 2nd transfer (n=2)	12.5 / 12.5	25 / 25	12.5 / 12.5				D
Distribution after 3rd transfer (n=3)	6.25 / 6.25	18.75 / 18.75	18.75 / 18.75	6.25 / 6.25			E
Distribution after 4th transfer (n=4)	3.125 / 3.125	12.5 / 12.5	18.75 / 18.75	12.5 / 12.5	3.125 / 3.125		F
Total amount mg in each container	6.25	25.0	37.5	25.0	6.25		
Fractions of solute in each container	0.0625	0.25	0.375	0.25	0.0625		

Buffer solution / CHCl$_3$

Figure 36-1. Theoretical distribution of solute after varying numbers of transfer.

Distribution curves may be prepared from hypothetical data using Equations 3 and 4 or from a computer program using these equations. Figure 36-3 illustrates a series of curves for a solute in which $K = 1.0$ and $R = 1.0$ following 8, 32, and 128 transfers. It is interesting to observe that as the number of transfers increases, the amplitude of the curve decreases and the solute spreads through more and more tubes. At first thought, this would seem undesirable, but the significant point is that the fraction of vessels containing solute after 128 transfers is now much less than after 10 transfers.

Therefore, two solutes with different but similar K values can be separated in 128 transfers because each solute occupies a smaller fraction of total tubes. If this separation were attempted with 10 to 20 transfers, both solutes would occupy nearly all of the tubes and no separation would be obtained.

Figure 36-4 illustrates the distribution patterns obtained in a 16-transfer experiment for solutes having distribution coefficients that differ by one order of magnitude. Under no circumstances can a separation be obtained if the distribution coefficients of the solutes are equal.

The procedure of operation that has been considered thus far is known as the *fundamental procedure.* Here, the solute is distributed through a specified number of tubes and nothing is withdrawn from the system until the entire operation is completed. Then the tube contents are withdrawn and analyzed for the purpose of determining solute concentrations, or the solutes are withdrawn simply for the purpose of isolating them from a mixture.

Another procedure of operation that is of interest primarily due to its analogy to elution chromatography is known as *end withdrawal.* In this operation the fundamental procedure is followed for a predetermined number of transfers as previously described. Then the upper phase only of the last tube in the train is collected. All other upper phases are advanced to the next tube in succession and after equilibration the upper phase of the last tube, n, is again collected.

This process is continued until all upper phases have passed through n tubes containing lower phase. In elution chromatography the analogy is similar. However, fresh upper phase is added continuously to the first *tube* (called a *plate* in elution chromatography) until only upper phase is eluted from the column.

In summary, the degree of separation of two or more solutes using CCD depends upon the distribution coefficients of the solutes, nature and volume of the solvents used, and number of transfers taken.

CENTRIFUGATION

A large number of separations may be accomplished with the centrifuge. This apparatus consists essentially of a container in which a mixture of solid and liquid, or of two liquids, is rotated at high speeds so that the mixture is separated into its constituent parts by the action of centrifugal force. A solid or liquid, mixed with a liquid of lesser density, may be separated be-
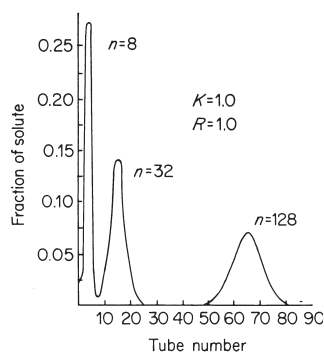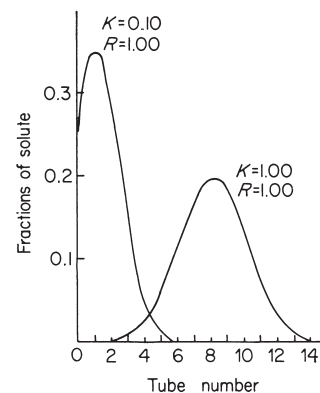
Figure 36-4. Distribution of two solutes with different K values.

cause the substance of higher specific gravity is thrown outward with greater force—it will be impelled to the bottom of the container, leaving a clear supernatant layer of pure liquid.

Centrifugation is useful particularly when separation by ordinary filtration is difficult, as in separating a highly viscous mixture. Separations may be accomplished more rapidly in a centrifuge than under the action of gravity. In addition, the degree of separation that is attainable may be greater because the forces available are of a far higher order of magnitude. The centrifuge has become a valuable analytical tool, particularly in biochemical and microbiological research. It has wide application in pharmaceutical laboratories and its use as a means of predicting emulsion stability has been suggested.

Two basic types of centrifuges are available: *sedimentation* and *filtration*. The *sedimentation type* of centrifuge depends on differences in the densities of the two or more phases comprising the mixture. This instrument is capable of separating both solid–liquid and liquid–liquid mixtures. *Filtration centrifuges,* however, are limited to the separation of solid–liquid mixtures only.

Sedimentation Centrifuges

The design of the bottle centrifuge and the disc centrifuge are based on the sedimentation principle (ie, separation by density difference).

BOTTLE CENTRIFUGE

The bottle centrifuge, which consists of a vertical spindle that rotates the containers in a horizontal plane, commonly is used to separate materials of different densities. Separation in a centrifugal field is brought about because denser particles in a mixture require greater forces to hold them in a circular path of a given radius than do lighter particles. Thus, the lighter particles are displaced toward the axis of the centrifuge by the heavier particles. During the centrifugation of blood, for example, a speed of 3000 rpm is required to separate blood corpuscles from serum. If the radius of the centrifuge is assumed to be 10 cm, the acceleration, a, acting on a particle can be approximated to be 10^6 cm/sec^2; or about 1000 times the acceleration due to gravity, g

$$a = 4\pi^2 N^2 r = \frac{4(3.14)^2(3000)^2(10)}{3600} = 10^6 \text{ cm/sec}^2$$

N = revolutions/sec; r = radius in cm

$$\frac{10^6 \text{ cm/sec}^2}{10^3 \text{ cm/sec}^2} = 100 \ (g)$$

10^3 cm/sec^2 = approximate acceleration due to gravity

Figure 36-3. Distribution of solute after varying number of transfers.

Under these conditions, the blood corpuscles eventually migrate under the influence of centrifugal force to the tip of the centrifuge tube.

The separation of particles in a liquid medium also depends on the nature of the medium. A solid particle settling under the influence of acceleration due to gravity in a liquid phase accelerates until a constant terminal velocity is reached. The terminal velocity is known as the settling velocity of the particle and is described mathematically by Stokes' Law. It can be shown that Stokes' Law can be extended to those cases where settling takes place in a centrifugal field,

$$v_s = v_g \frac{\omega^2 r}{g} \qquad (5)$$

where v_s is the settling velocity of a particle in a centrifugal field, v_g is the settling velocity of a particle in a gravitational field (Stokes' Law), ω is the angular velocity of the particle in the settling zone, and r is the radius at which the settling velocity is determined.

Consider a solid particle at an initial position in a liquid medium and a distance r from the axis of rotation. Under these conditions,

$$v_s = dr/dt \qquad (6)$$

Substituting Equation 6 into Equation 5 gives

$$dr/dt = v_g \frac{\omega^2 r}{g} \qquad (7)$$

Rearranging and integrating between limits gives

$$\int_r^{rc} \frac{dr}{r} = \int_0^t v_g \frac{\omega^2 r}{g} \, dt \qquad (8)$$

$$\ln \frac{r_c}{r} = v_g \frac{\omega^2 t}{g} \qquad (9)$$

where r_c is the distance between the surface of the sedimented cake in the tip of the tube and the axis of rotation, and t is the time during which the particle is subjected to centrifugal acceleration while the particle travels the distance from r to r_c. Equation 9 shows that if centrifuging conditions for a given suspension are to be compared in different centrifuges, the speed, bottle size, centrifuge dimensions, and centrifuging time must be taken into consideration. Lavanchy and Keith[3] describe mathematical approaches that should be taken for this purpose.

ULTRACENTRIFUGE

When extremely fine solid matter must be separated from a liquid, such as in colloid or biological research, the ultracentrifuge is employed. In this instrument a relatively small rotor is operated at speeds exceeding 100,000 rpm and forces up to one million times gravity are exerted. High speeds are attained with air or oil turbines and bearings lubricated with a film of compressed air. Friction heat may be minimized by the use of high vacuum.

By placing the samples in specially constructed cells and spinning them in the ultracentrifuge, it is possible to separate the dispersed phase from the continuous phase rather rapidly. To aid the investigator, optical attachments may be employed to photograph the settling while the centrifuge is in operation.

Only small batches of material can be handled in these instruments during a single run. Ultracentrifuges are employed in the determination of particle size and molecular weight of polymeric and other high-molecular-weight materials such as proteins and nucleic acids by direct or indirect observation of the rate of separation of particles in solution or suspension.

Filtration Methods

The filtration centrifuge is restricted to the separation of solid–liquid mixtures. It is similar in principle to the sedimen-

tation type, but rather than containers it possesses a porous wall through which the liquid phase may pass but upon which the solid phase is retained. Analogous to filtration, this process requires consideration of the flow of liquid through the solid bed that accumulates on the porous plate.

FILTRATION

Filtration is the process of separating liquids from solids with the purpose of obtaining optically transparent liquids. This is accomplished by the intervention of a porous substance, called the *filter* or the *filtering medium*. The liquid that has passed through the filter is called the *filtrate*.

Mathematics of Filtration

In 1842 Poiseuille proposed a relationship for streamlined flow of liquids under pressure through capillaries. This equation in its simplified form is represented by

$$V = \frac{\pi \Delta p r^4}{8 L \eta}$$

where V = flow velocity, r = flow capillary radius, L = capillary length, η = viscosity of the fluid, and Δp = pressure differential at the two ends of the capillary.

The modified Poiseuille equation has been shown to be valid for liquid flow through sand, glass beads, and various porous media. It represents the foundation for all mathematical models of filtration that where developed subsequently. Of critical importance in this equation is the powerful effect of capillary radius; ie, by reducing it to 1/8 its size, the pressure differential must be increased more than 4000 times in order to obtain the same flow velocity, all other factors remaining constant.

On the basis of the Poiseuille formula, the Kozeny-Carman relationship was established. This may be expressed as

$$V = \left[\frac{e^3}{KS^2(1-e)^2} \right] \left[\frac{A \Delta p g}{\eta L} \right]$$

where A = cross-sectional area of porous bed (filter medium), e = porosity of bed, S = surface area of medium, K = constant, and the remaining symbols are the same as in the Poiseuille equation.

The Kozeny-Carman relationship, like Poiseuille's law states that the rate of flow is directly proportional to the pressure drop across the medium and to the area of the bed, and inversely proportional to the viscosity of the liquid and the thickness of the bed. To characterize the material composing the bed, two new quantities, e and S, are introduced, replacing capillary radius.

The use of a nondefinite constant K, rather than the definite constant in Poiseuille's equation, $\pi/8$, offers greater utility in the use of this equation in accounting for the geometry of the medium. The constant, K, generally ranges in value from 3 to 6. The Kozeny-Carman equation finds its greatest limitation in complex systems such as filter paper, but provides excellent correlation in filter beds composed of porous material.

In applying Poiseuille's law to filtration processes, one must recognize the capillaries found in the filter bed are highly irregular and nonuniform. Therefore, if the length of a capillary is taken as the thickness of the bed or medium and the correction factor for the radius is applied, the flow rate is more closely approximated. These factors have been taken into account in the formulation of the Darcy equation

$$V = \frac{k \Delta p}{L \eta}$$

where k is the permeability coefficient and depends on the nature of the precipitate to be filtered and the filter medium itself.

Computer-assisted design of microfiltration systems are underway.[4] This technique is used to design an optimum filtration system from actual filtration data, thereby predicting its performance with any given fluid.

In considering the nature of the precipitate, it is known that large particles are easier to filter than are small particles because of the tendency of the latter to enter into and occlude the pores of the bed, thus hindering the passage of the filtrate. In addition, the buildup of small particles on the filter tends to form a nonporous, densely packed bed that also resists passage of the filtrate.

Filtering Media

The filtering medium, whether a filter paper, synthetic fiber, or porous bed of glass, sand, or stone, is composed of countless channels that impart porosity to the medium. Almost without exception these channels or pores are nonuniform and possess a rather tortuous nature.

The mechanism of filtration basically involves a two-step process:

1. The filter medium itself resists the flow of solid material while permitting the passage of liquid.
2. During the course of the filtration the suspended, solid material builds up on the filter medium and thereby forms a *filter bed*, which acts as a second, and often more efficient, filter medium.

The ability of a filter medium to eliminate solid matter from a liquid is termed *retention*. It must be borne in mind that the filtration process must compromise retention with filtration rate, the speed at which the purified liquid (the filtrate) is recovered. To illustrate this point, it will be noted that a slab of marble will most effectively retain the solid material contained in a suspension; unfortunately, it would require a few centuries to collect the purified filtrate.

Both the retentive ability of a filter medium and filtration rate of a liquid through the medium depend on the porosity of the medium. Each factor, however, is influenced significantly by the viscosity of the liquid, the proportion of solid matter in the liquid, and the size, shape, and physical nature of the suspended solids.

The flow of a liquid through a filter bed follows the same basic rules that govern the flow of any liquid through a medium offering resistance. The *flow rate* through the medium will vary directly with the area of the medium, as well as the pressure drop or driving force across the bed.

$$\text{Rate of flow} \propto \frac{(\text{driving force})(\text{cross-sectional area})}{\text{resistance}}$$

The flow rate is retarded by the viscosity of the liquid being filtered and by any obstruction to flow. These obstructions include the resistance of the filter medium itself and the second filter bed or *filter cake* that builds up on the medium at a rate dependent on the solids content of the liquid. The resistance offered by the medium itself will not vary significantly during the filtration process. It depends on the thickness of the medium as well as its porosity. The resistance of the filter cake, on the other hand, is not constant and generally increases continuously during the operation. The resistance offered by the cake depends both on its thickness and physical nature. The thickness of the cake is dictated by the amount of filtrate passing through the filter and on the solids content of the liquid. The physical nature of the cake—whether it is loose, compacted, coarse, fine, granular, or gelatinous—determines whether or not it will readily allow the flow of liquid.

FILTER PAPER

Filter paper most frequently is employed in clarification processes required of the pharmacy practitioner. Only high-quality filter paper should be used to ensure maximum filtering effi-

ciency. When possible the first few milliliters of filtrate should be discarded to eliminate (insofar as possible) contamination of the pharmaceutical product by free fibers associated with most filter paper. This is especially true in the preparation of ophthalmic solutions.

MEMBRANE FILTERS

Membrane filter media are produced from pure cellulose, cellulose derivatives, and polymeric materials. All have an extremely uniform micropore structure as well as an exceptionally smooth surface. The integral structure contains no fibers or particles that can work loose and contaminate a filtrate. This is a particular advantage in the filtration of ophthalmic solutions. The presence of these fibers is difficult to prevent when using many other filter media, including paper filters.

The efficiency of membrane filters is due to the uniform pore system that functions like a highly effective sieve. The pore size, of different types of these filters, ranges from 10 nm to 10 μm. All particles in liquids or gases that are larger than the pore of a given filter are retained on the surface. The thickness of these membrane filters ranges from 50 to 200 μm.

The pores that penetrate these filters pass directly through the entire thickness of the membrane, with a minimum of crosslinkage. Porosity or pore volume is estimated as 80% of the total fiber volume. The high porosity of these filters, coupled with the *straight-through* configuration of the pores, results in flow rates through membrane filters that are at least 40 times faster than flow rates through conventional filter media that possess the same particle size retention capabilities.

Major producers of these filters include the *Millipore Filter Corp*, Bedford, MA; *Gelman Instrument Corp*, Ann Arbor, MI; *Pall Corp*, Glen Cove, NY; *Nuclepore Corp*, Pleasanton, CA; and *Carl Schleicher & Schuell Co*, Keene, NH. The membrane filters are available as circular discs of varying diameter. Different types are available for use in the filtration of either aqueous or nonaqueous liquids. The discs generally are used in conjunction with specialized holders of either metal or glass composition. With small volumes (ie, less than 500 mL), solutions usually are filtered using vacuum techniques. Larger volumes require filtration under pressure provided by an inert gas such as nitrogen.

In addition to their obvious utility in routine filtration processes on both a laboratory and industrial scale, these filters have been used for a wide range of purposes, including chemical analysis, microbiological analysis, and bacterial filtration. The latter process provides an economical and rapid method for sterilizing heat-labile material (see Chapter 40).

OTHER FILTERING MEDIA

Many devices have been advanced to replace filter paper, which has many disadvantages, particularly for large-scale operations. A great many variations of filtering processes, each designed to fit the needs of special cases, are found in the modern pharmaceutical laboratory. The filter press, the centrifugal filter, the vacuum filter, sand-bed filter, charcoal filter, paper-pulp filter, and porous porcelain filter are all examples of specialized filtration methods. Each one of these possesses some advantageous quality, and it is the experience of the laboratory operators that guides them in their selection of appropriate filtering devices. Reference is made later in the text to many of these special-scale filters.

However, it would not be inappropriate to refer briefly to special filtering devices that may be useful in the prescription or research laboratory.

Cotton Filters—A small pledget of absorbent cotton, loosely inserted in the neck of a funnel, adequately serves to remove large particles of extraneous material from a clear liquid. Although this properly might be termed colation, the cotton also can be used to serve as a fairly efficient filter. It is sometimes necessary to return the liquid a number of times to secure perfect transparency. This should be remembered in filtering

ophthalmic solutions through cotton, because small detached filaments are carried through on initial filtration.

Glass-Wool Filters—When solutions of highly reactive chemicals, such as strong acids, are to be filtered, filter paper cannot be used. In its place glass wool may be used just as one uses absorbent cotton for filtering. This material is resistant to ordinary chemical action, and when properly packed into the neck of a funnel it constitutes a very effective filtering medium.

Sintered-Glass Filters—These filters have as the filtering medium a flat or convex plate consisting of particles of Jena glass powdered and sifted to produce granules of uniform size that are molded together. The plates can be fused into a glass apparatus of any required shape (Fig 36-5). These filters vary in porosity, depending on the size of the granules used in the plate. They are very useful in the filtration of solutions such as those intended for parenteral injection. A vacuum attachment is necessary to facilitate the passage of the liquid through the filter plate (see Chapter 40).

Funnels

Funnels are conical-shaped utensils intended to facilitate the pouring of liquids into narrow-mouthed vessels. They also are used widely in pharmacy for supporting filter media. Funnels may be made of glass, polyethylene, metal, or any other material that serves a specific purpose. The community pharmacist will find the glass funnel to be quite adequate for all processes of clarification in prescription practice.

Most funnels used by the pharmacy practitioner are conical in shape and may be fluted, grooved, or ribbed for the purpose of facilitating the downward flow of the filtrate.

The *Büchner* type of funnel is used today largely in pharmaceutical laboratories. A piece of round filter paper is laid on the perforated porcelain diaphragm and the filtration conducted. This funnel is especially applicable to vacuum filtration (see the discussion, *Vacuum Filtration*).

FILTRATION OF VOLATILE LIQUIDS

It is evident that the ordinary methods of filtering liquids will not be practical for very volatile liquids because of the loss through evaporation, and the liability to explosion in the case of flammable volatile liquids. Funnels must be covered, the receiving vessel closed, and provision made for the escape of the confined air in the receiving vessel. The following method is quite useful. A rubber cover, perforated to admit a tube, is placed on top of the funnel; connection between the bottle and funnel is effected as shown in Figure 36-6.

AIDS TO FILTRATION

It has long been known that addition of an insoluble adsorbent powder to a liquid prior to its filtration greatly increases the
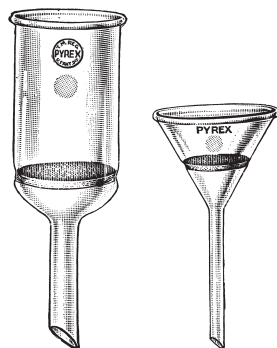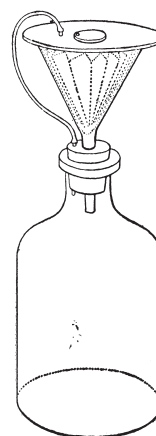

Figure 36-5. Sintered-glass filters.

Figure 36-6. Filtration of volatile liquids.

efficiency of the process. Purified talc, siliceous earth (kieselguhr), clays, charcoal, paper pulp, chalk, magnesium carbonate, bentonite, silica gel, and others have been used for this purpose.

It must not be overlooked, however, that powdered substances employed for such purposes must be insoluble and inert, so not all of those in the foregoing list are applicable for general filtration.

Talc is nonadsorbent to materials in solution and is a chemically inert medium for filtering any liquid, provided it has been purified for this purpose and it is not the impalpably fine variety that will pass through the filter paper.

Kieselguhr is almost pure silica (SiO_2). It is as applicable as talc for general filtration purposes, with no danger of removing active constituents by adsorption.

Siliceous earths or clays, such as fuller's earth or kaolin in the hydrated form which is produced when they are brought into contact with aqueous liquids, are safe for general use only in filtering fixed oils. Liquids containing coloring matter or alkaloidal principles must not be filtered through these media, for adsorption of both color and alkaloids occurs and the filtrate is altered in comparison.

Charcoals, as a rule, possess adsorptive properties not only toward color but for many active constituents of medicinal preparations, such as alkaloids and glycosides. Consequently, charcoal should never be used as a filtering medium unless the removal of such constituents is desirable.

Chalk and *magnesium carbonate* readily react with acids and possess a finite solubility in water and aqueous fluids, with the production of alkalinity in the filtrate. This is particularly true of magnesium carbonate; the degree of alkalinity imparted to the filtrate is sufficiently great to cause precipitation of alkaloids. Either of these media, when added to an alkaloidal preparation prior to filtration, will precipitate and remove all of the alkaloidal constituents. Neither is suitable for general use.

RAPID FILTERING APPARATUS

Much attention has been given to methods for increasing the rapidity of filtration. This may be accomplished by applying pressure on the filter or by creating a vacuum in the receiving vessel.

VACUUM FILTRATION

One of the first practical efforts made to create a vacuum to aid filtration was by means of the Bunsen pump. Its action depends on the principle that a column of water descending through a tube from a height is capable of carrying with it the air contained in a lateral tube, if the latter is placed properly. This form of aspirator is practicable where water pressure is available.

Pumps Acting by Water Pressure—The various aspirator or vacuum pumps that operate under the influence of water pressure are all based on the same principle. The following are selected for illustration from the great variety in use. Figure 36-7 shows Chapman's vacuum pump. Valve *a* prevents the water from flowing into the bottle which carries the filter when the pressure of water ceases or is reduced.

On a larger scale, the vacuum for filtration is produced by one of the many types of vacuum pumps now available. The pump should be protected from vapors by placing a suitable vapor trap between the filter unit and the pump. The trap usually is cooled to very low temperatures by means of dry ice and acetone when very high vacuum is needed.

In assembling a filtering apparatus using the vacuum principle, it is necessary that there be no leaks in the connections from the filter to the aspirator. If filter paper is used in connection therewith, a plainly folded paper must be used and its tip must be protected against breakage by reinforcing it with a filter paper support or some other device. A Büchner filter also may be used, employing a specially strong filter paper.

In analytical work it is customary to use the Gooch crucible and flask (Fig 36-8) for rapid filtration. The flask, of especially thick glass, is provided with a side tube that is connected to a water aspirator pump. The perforated crucible bottom is converted into a filter bed of the required thickness by means of a filter mat placed over the perforations in the porcelain base.

FILTRATION UNDER PRESSURE

Figure 36-9 illustrates a sectional drawing of a plate-and-frame filter press. The material to be filtered enters the apparatus under pressure through a pipe at the bottom and is forced into one of the many chambers. A filter cloth is positioned on both sides of each chamber. As the material passes through the filtering cloth, solids remain behind in the chamber and the clear filtrate passes through and out of an opening located on top of the apparatus.

Rotary-drum vacuum filters are used widely in the pharmaceutical industry, especially in the preparation of antibiotics by the fermentation process. In this type of filtration a perforated drum, wrapped with a cloth or other suitable substance holding a filter medium, is immersed partially in a tank holding the material to be filtered (Fig 36-10).
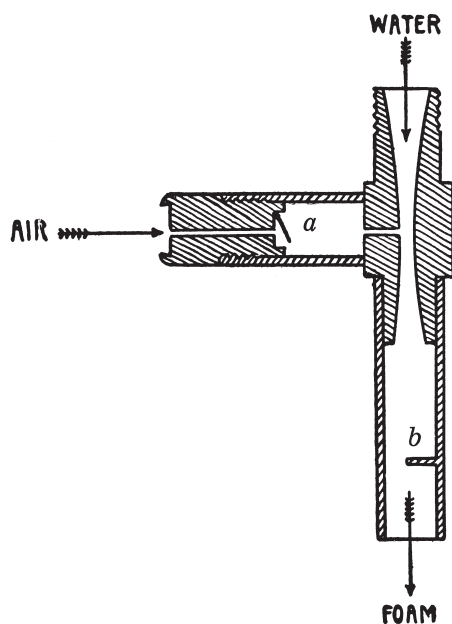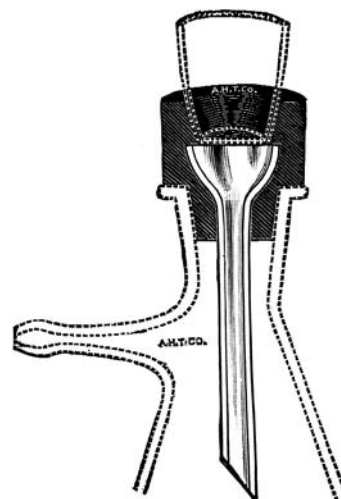

Figure 36-8. Gooch crucible arranged for vacuum filtration (courtesy, Thomas).

The drum is rotated through the slurry of material and a vacuum within the drum draws the material into and through the filter medium. During this step of the process, the filtrate is taken into the drum and collected, while the solid material remains deposited on the outer surface of the drum. This material is then removed by a scraper in the last step of the operating cycle, just before the rotating drum repeats another cycle.

CLARIFICATION AND DECOLORATION

Clarification

Clarification is the process by which finely divided solids and colloidal materials are separated from liquids without the use of filters. The process is employed to remove suspended oil from aqueous solutions, such as aromatic waters, and for the removal of undesirable solids that interfere with the transparency of such natural products as honey and fruit juices.

Clarification generally is resorted to when the contaminating material is finely subdivided, amorphous, or colloidal in nature and tends to plug a filtration medium rapidly. A number of methods are available to handle this difficult problem.

When the solids are not of a granular or free-filtering nature, it may be possible to improve the characteristics of the suspended solids. This may involve varying the temperature or pH of the medium. When a viscid liquid is heated, its viscosity and specific gravity are decreased and particles that are suspended in it will separate. Those particles that are more dense than the liquid will fall to the bottom, while those that are less dense will rise to the surface. In the latter case the minute bubbles of steam formed in the heating process become enveloped in the viscid particles, rise through their buoyancy, and a scum is formed that may be separated readily.

The dewaxing of oils at a reduced temperature offers a further example of the possibilities of contaminant modification. Oil that is chilled rapidly often produces an amorphous wax that will plug a straining medium. Slow chilling, on the other hand, produces a wax with a more crystalline nature, which has good filtration characteristics.

The simplest method of clarification, although not always feasible, is gravitational sedimentation. This method involves the least amount of labor and expense and is used frequently, particularly on a large scale, when haste is unnecessary. The deposit formed is called a *sediment* or *sludge*. These terms are not synonymous with *precipitate*. A sediment is solid matter separated merely by the action of gravity from a liquid in which it has been suspended. A precipitate, on the other hand,
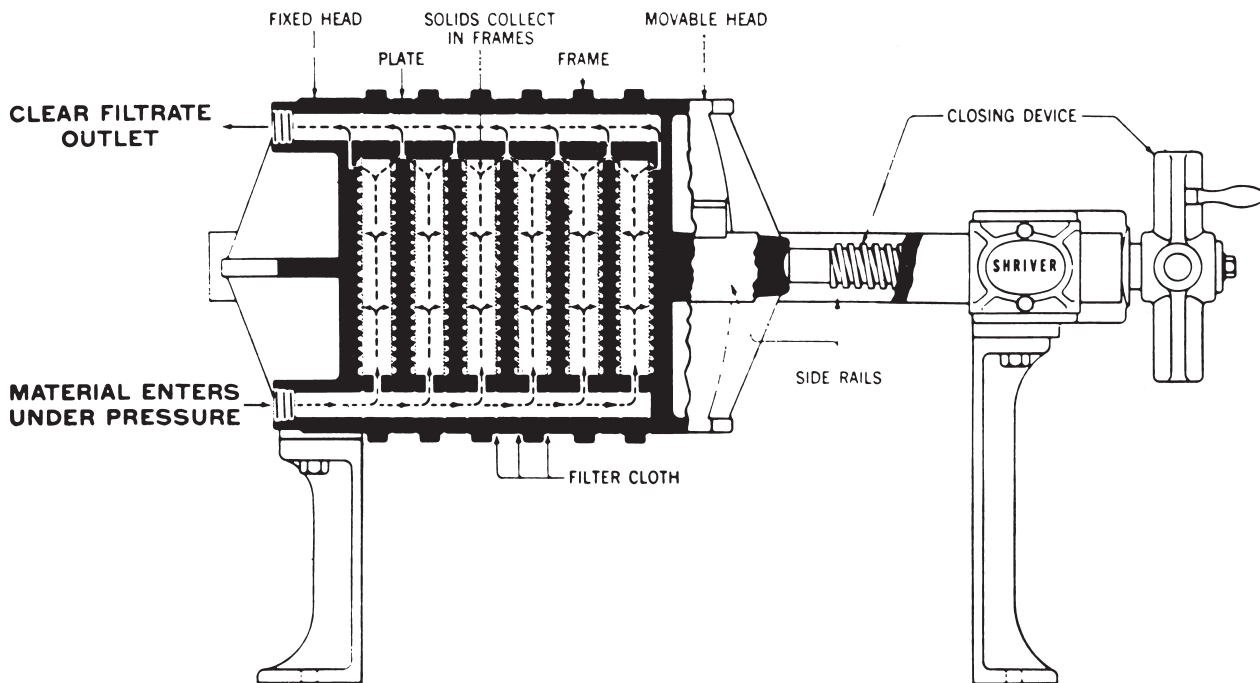

Figure 36-7. Chapman's pump.

Figure 36-9. A plate-and-frame filter press (courtesy, Shriver).

is solid matter separated from a previously clear solution by physical or chemical change. Fixed oils usually are clarified by gravitational sedimentation. In vegetable oils the sediment consists principally of albuminous and gummy substances, cellular tissue, and water, all of which have been separated with the oil during the expression process.

The clarification process generally is carried out by adding a clarifying agent such as paper, pulp, talc, infusorial earth, as well as a number of other materials to the turbid liquid. These agents usually act to reduce turbidity by physical adsorption of the contaminating material, although a large number of specific, physical-chemical coagulants also are in use. After the addition of the clarifying agent, the mixture is agitated and the agents, along with the adsorbed impurities, are removed by filtration or any other suitable means. Albumin and gelatin are examples of clarifying agents obtained from natural sources.
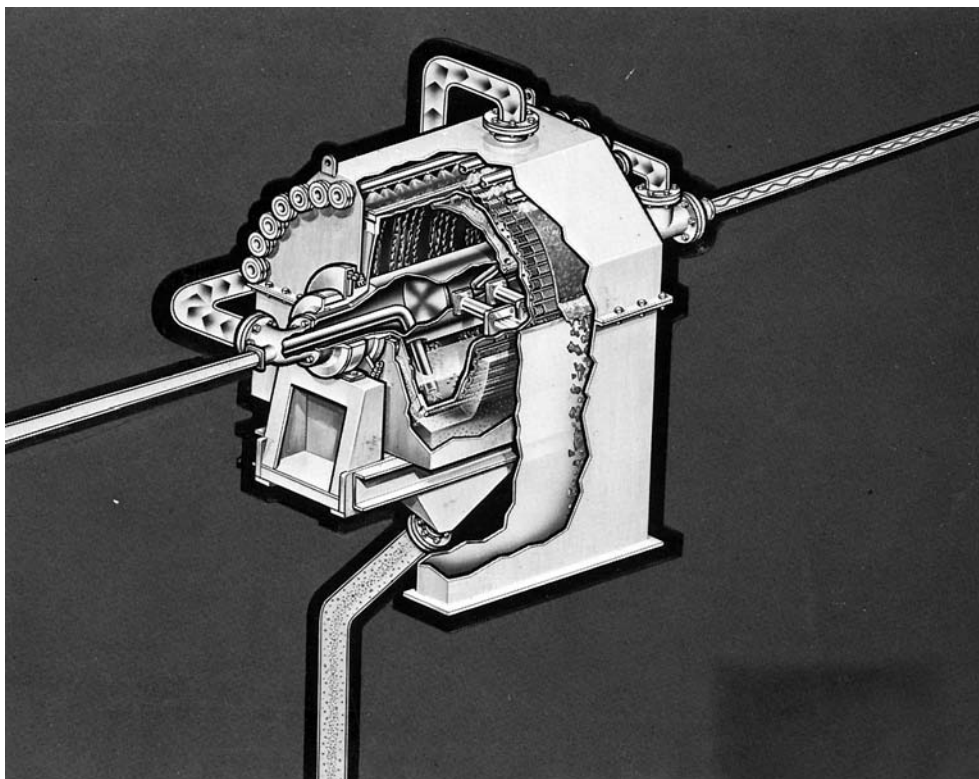

Figure 36-10. Rotary filter (courtesy, Bird Machine).

Substances of a synthetic nature, such as polyamines, also are used for this purpose.

Decoloration

Decoloration, or decolorization as it sometimes is called, is the process of depriving solutions of color by use of an appropriate adsorptive medium. In many respects it is closely related to the clarification process. Decoloration is used for removal of coloring matter from a number of raw materials, both natural and synthetic, and from many finished products. Animal charcoal (also called bone black), wood charcoal, or activated charcoal frequently are used as decolorizing agents. Clays such as bentonite, kaolin, and fuller's earth also are used for this purpose.

LOTION, DECANTATION, AND COLATION

Lotion

Lotion (displacement washing) is the process by which soluble impurities are removed from insoluble material by the addition of a suitable washing solvent. The wash liquid usually is separated from the purified solid by decantation or filtration. An expedient method of adding the washing solvent to the solid in a fine, controlled spray is by the use of wash bottles.

CONTINUOUS WASHING

The use of the wash bottle is limited to small operations. A simple method of automatically supplying the wash liquid in larger quantities is shown in Figure 36-11. This requires attention from the operator only at the beginning of the operation. The inverted bottle containing the washing solvent is furnished with a perforated stopper and a short glass tube. All that is necessary is to fill the bottle and adjust it over the funnel so that the end of the tube is at the height at which the level of liquid in the funnel is to be maintained. When the bottle is tilted slightly (if the tube selected is not too narrow in diameter), the liquid runs into the funnel until it rises to the orifice of the tube, whereupon the flow ceases. As the liquid gradually passes through the solid substance in the funnel, the level falls below the orifice, bubbles of air pass through the tube into the bottle, the liquid once more flows, and the operation continues until the upper bottle is empty. Many elaborate methods of continuous washing have been suggested, but the simple apparatus just described is quite satisfactory if a tube of proper diameter has been selected, one of such size that the force of capillary attraction will not be strong enough to prevent the passage of air.
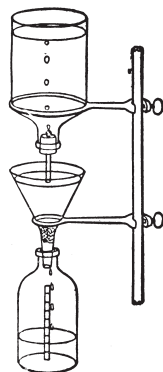

Figure 36-11. Continuous washing.

Decantation

The simplest method available for the separation of a solid from its soluble impurities is the technique of decantation. This method involves washing and subsequent agitation of the solid with an appropriate solvent, allowing the solid to settle and removing the supernatant solvent. These three steps are repeated as often as required to attain the desired purity of the solid. This method also is applicable to the simple separation of solids and liquids, such as after precipitation of a material from a mother liquor. Decantation provides an effective method for washing magmas and other gelatinous products.

Some degree of skill is required to decant liquids effectively. It is most convenient to decant from a lipped vessel that is not filled to capacity. In addition, the use of a stirring rod is suggested as a guide to steady the hand of the operator.

Colation

Colation or straining (from Latin *colare,* to strain) is the process of separating a solid from a fluid by pouring the mixture on a cloth or porous substance that will permit the fluid to pass through, but will retain the solid. This operation frequently is used for separating sediment or mechanical impurities of various kinds from liquids.

Colation should not be considered as a separate process but simply as a crude form of filtration, with larger pores in the straining medium than usually are employed for filtration.

The essential apparatus is a straining medium and a strainer support or frame. The straining medium is usually a cloth material such as flannel, muslin, wool, or cheesecloth. The material should be colorless and washed before use. Fabrics, particularly those of cotton, usually are treated or impregnated with a material called *sizing* to improve their appearance and quality for certain purposes; however, for use as a strainer, the fabric must be free of sizing because it causes contamination. Many different substances are used for sizing, some being soluble in cold water, others only in hot water. Thus, the proper method for their removal is to soak the fabric for a few hours in cold distilled water, rinse thoroughly; then cover with distilled water, boil for a few minutes, and rinse well in distilled water to remove the last traces of the gelatin, albumin, glue, or starch that may have been present in the sizing.

EXPRESSION

Expression is a process of *forcibly* separating liquids from solids. A number of mechanical principles have been recognized in the operation of expression, namely the use of the spiral twist press, the screw press, the roller press, the filter press, and the hydraulic press.

SPIRAL TWIST PRESS—The principle of this press is best and most practically illustrated in the usual process of manually expressing a substance contained in a cloth.

ROLLER PRESS—This is used for large-scale pressing of oily seeds, fatty substances, and so on. Care must be taken to apply the force gradually to the bag containing the material to be pressed, and not to use it on substances that will be corrosive to the rubber rollers.

HYDROSTATIC OR HYDRAULIC PRESS—Of the presses heretofore mentioned, each has some special advantage of use, but each also has some objectionable feature. The spiral twist is not powerful and its action is limited. The screw presses have friction with which to contend; the friction of a screw increases with the intensity of the pressure applied, and when a certain limit is reached all further force applied is wasted, and if continued may result in destruction of the press. The roller press is very limited in its action. Although the hydraulic press is expensive, after the first coat it is the most economical because the greatest power is obtained at the expense of the least labor. The principle of a hydraulic press is based on the fact that pressure exerted

upon an enclosed liquid is transmitted equally in all directions. Tremendous pressures can be developed with hydraulic presses.

PRECIPITATION

Precipitation is the process of separating solid particles from a previously clear liquid—a solution—by physical or chemical changes. The separated solid is termed a *precipitate;* the cause of precipitation is the *precipitant;* and the liquid that remains in the vessel above the precipitate is called the supernatant liquid.

In pharmacy, precipitation may be useful for many purposes. It provides a convenient method of obtaining solid substances in the form of fine particles, such as the precipitation of calcium carbonate (precipitated chalk). White Lotion is an example of a preparation prepared by precipitation, in this case by mixing aqueous solutions of zinc sulfate and sulfurated potash to form an insoluble, finely divided zinc sulfide, free sulfur, and various polysulfides.

One of the most important uses of precipitation is in the purification of solids. The process as applied to purification is termed *recrystallization.* The impure solid usually is dissolved in a suitable solvent at elevated temperatures. On cooling, the bulk of the impurities remain solubilized while the purified solid product precipitates. This procedure is repeated as many times as necessary, using a number of solvents if required.

SEPARATION OF IMMISCIBLE LIQUIDS

The separation of liquids that are mutually soluble usually is effected by distillation, if one or both of the liquids are volatile. The separation of liquids that are immiscible is generally a simpler process.

Separations of this kind are necessary in analytical procedures, manufacturing operations, distillation of volatile oils, and accidental contaminations and admixtures, and are usually best made using a separatory funnel. When very small amounts of liquids are floating on the surface of another liquid, separation is accomplished most easily by using a pipet, medicine dropper, or glass syringe with an attached needle.

FLORENTINE RECEIVER

The separation of volatile oils from the water that accompanies them during steam distillation is a very important part of their manufacturing process. Where the volatile oil is lighter than water, the principle shown in Figure 36-12 may be used. The oil and water collect in the glass receiver during distillation, the oil floating on the top, while the water ascends the bent tube from the bottom; further addition of distillate causes the water to overflow from the side tube. The reverse action is produced in the receiver for light or heavy oils (Fig 36-13), in which either a lighter or a heavier fraction may be collected continuously.
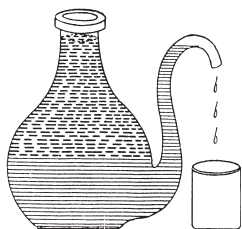

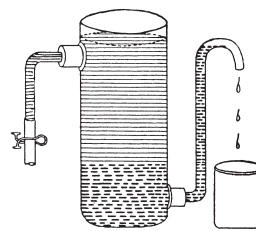**Figure 36-12.** Florentine receiver.

Figure 36-13. Receiver for light or heavy oils.

SPECIALIZED SEPARATION TECHNIQUES

Diffusion Phenomena

Diffusion is the spontaneous penetration of one substance into another under the potential of concentration gradient. Simply stated, material will tend to move from a region of higher concentration to one of lower concentration. The driving force or potential of such a process may be enhanced by the application of an electric field.

If the two regions of concentration noted are separated by a selective membrane, certain species will diffuse through the membrane, while other molecular species will be held back. When this selectivity is dictated by the porosity of the membrane, the process is termed *dialysis.* Dialysis is used principally for the separation of small molecules and ions contained in a mixture with colloidal material. The latter substances diffuse with difficulty or not at all. Materials such as gums, starch, albumin, and proteins fall into this colloidal, nondiffusible category.

The rate of diffusion across a semipermeable membrane is directly proportional to the concentration gradient between the two surfaces of the membrane and to the area of the membrane, but is inversely proportional to the membrane thickness. These factors are expressed in Fick's law of diffusion

$$\frac{dS}{dt} = \frac{kA(C_i - C_0)}{h}$$

where S is the amount of substance diffused at time t, k is a permeability constant, A is the membrane area, h is the membrane thickness, dS/dt is the diffusion rate, C_i is concentration on one side, and C_0 is concentration on the other side of the membrane.

Gel Filtration

The chromatography of cephalosporins in gel filtration chromatography has been demonstrated and shown to be important in the separation of high-molecular-weight impurities. The impurities frequently are associated with allergic responses in patients. This method has been demonstrated to serve as an excellent quality-control procedure for the impurities in cephalosporin preparations.[5]

Different types of Sephadex gels were used for separation. The study investigated various reagents necessary to perform the separation in an ultimate purification of the compound. The results indicated that optimization was capable of being done to separate the impurities from the active compound. The nature of the mobile phase, the ionic type, pH value, and molarity were important for the optimization.

A feasibility study of liposome separation that was undertaken to explore the use of size-exclusion chromatography, such as gel filtration of a large-scale process, demonstrated that it could separate liposomes from freeze-dried material in a chromosome preparation.[6] The chromatographic step was intended to improve the drug encapsulation by removing free (unincapsulated) drugs from external media. The selected stationary phase was G-50 Sephadex. The model drug used in the study was orciprenaline sulfate. The technique was able to produce a suitable size exclusion that efficiently removed the free drug from the liposome preparation.

In a study of liposomes loaded with calcitonin, it was necessary to observe the location of the protein to protect it from enzymatic digestion.[7] The analysis of the liposome produced from this protein was extracted using suitable gel separation of the liposome mixture to ensure the location of the protein within the system. It established the stability and the ultimate formation of the liposome product. This ensured the appropriate loading of the protein within the liposome product.

A process for purifying bovine pancreatic glucagon as a byproduct of insulin production was described.[8] The glucagon, containing supernatant from the alkaline crystalline crystallization of insulin, was precipitated using ammonium sulfate and isoelectric precipitation. The precipitate containing glucagon then was purified by ion-exchange chromatography on Q-Sepharose FF gel filtration on Sephadex G-25 and ion-exchange chromatography on S-Sepharose FF. Successful yields were obtained using this technique, which was successful because of the gel filtration procedure.

A report was presented on the characterization of adenosine receptors in porcine striatal membranes and their solubilization by detergent digitonin.[9] Once the drug was solubilized, the material was bound to sites after the removal of receptors from the lipid environment. Gel filtration on Superdex 200 accomplished the separation into appropriate molecular weights. Suitable purification was achieved by this means.

In another report of the use of gel filtration, the expression and purification of human gamma-glutamylcysteine synthetase were studied.[10] Specific proteins and polypeptides were isolated and their amounts characterized by the use of Superdex 200 along with ATP-affinity resins. Cyclosporin A has potential for wide clinical use, limited only by the very narrow therapeutic index.[11] Potentiation of its clinical efficacy is thus very desirable. Preliminary data had indicated that the mixture of cyclosporin A, with hyaluronate, could increase its efficiency. In this study, it was found that cyclosporin A could reduce the hypersensitivity in test animals when administered along with hyaluronate. To demonstrate the association of this mixture, gel filtration was required, which showed the protection of the molecule from being bound to red blood cells. This association would improve the clinical response and was proven only by the use of gel filtration.

Ultrafiltration

A new type of membrane coating has been developed for osmotic delivery that offers significant advantages over membrane coatings used in conventional osmotic tablets.[12] This coating has an asymmetric structure similar to the asymmetric membranes made for reverse osmosis membrane or ultrafiltration. The study demonstrated clearly how the porous membranes could work as a thin outer skin of a dosage form. The permeability of the coating to water can be controlled by the membrane structure, whose principles were derived from ultrafiltration principles. A porosimetric technique for verifying the integrity of virus-retentive membranes, which can be validated, has been studied.[13] This integrity test of filtration processes was specifically designed for and is useful for post-use membrane integrity testing.

Reverse Osmosis

As reverse osmosis (Fig 36-14) is used it is necessary to evaluate which new composite reverse osmosis membranes were developed with significant improved performance over commercially available conventional composite membranes. The ESPA membrane chemistry provides a high flux at low operating pressure while maintaining a very good salt and organic rejection. The membranes have been demonstrated to operate for several years. Appropriate transmission and field emission electron micrographs of the membrane demonstrated the struc-
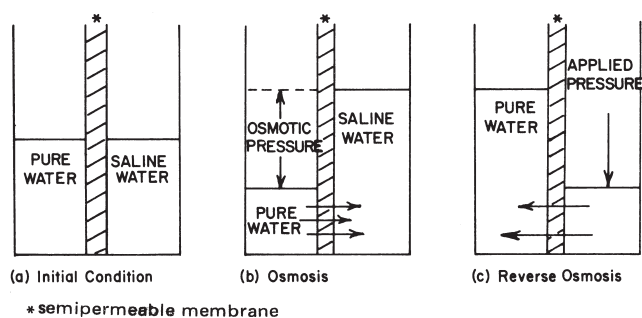

Figure 36-14. Principles of reverse osmosis.

ture of the membrane skin layer is the reason for the improved performance. This surface charge of the various membranes was demonstrated qualitatively using zeta-potential measurements. The newer membranes had a low surface charge and operated at a low pressure. In an effort to further improve the available reverse osmosis water-treatment membranes, other studies have been conducted over the past several years to evaluate specific ultra-low-pressure membranes. Very little information was available to the industry. It is possible to design membranes with 30% increase in productivity over conventional membranes. These improvements are particularly important to multistage systems for water purification. Recommendations have been made by many to improve the systems by using ultra-low-pressure membranes.

REFERENCES

1. Craig LC, Craig D. In Weissberger A. *Technique of Organic Chemistry,* vol 3, pt 1, 2nd ed. New York: Interscience, 1956, chap 2.
2. Rogers LB. In: Kolthoff IM, Elving PJ. *Treatise on Analytical Chemistry,* vol 2, pt 1. New York: Interscience, 1961, chap 22.
3. Lavanchy AC, Keith FW In: *Kirk-Othmer Encyclopedia of Chemical Technology,* vol 5, 3rd ed. NY: Interscience, 1991, p 194.
4. Weyand J. In: Shoemaker W. *What the Filterman Needs to Know about Filtration.* AIChE Symposium Series, no 171, vol 73. New York: American Institute of Chemical Engineers, 1977.
5. Changyin H, et al. *J Pharm Biomed Anal* 1994; 12:533.
6. Nemuri S, Rhodes C. *Pharm Acta Helv* 1994; 69:107.
7. Arien A, et al. *Pharm Res* 1995; 12:1289.
8. Andrade A, et al. *J Med Biol Res* 1997; 30:1421.
9. Costa B, et al. *Neurochem Int* 1998; 32:121.
10. Misra I, Griffith O. *Prot Express Purif* 1998; 13:268.
11. Gowland G. *Int J Immunotor* 1998; 14:1.
12. Herbig S, et al. *J Contr Rel* 1995; 25:127.
13. Phillips M, Diheo A. *Biologicals* 1996; 24:243.

BIBLIOGRAPHY

Curling JM. *J Parenteral Sci Technol* 1982; 36:59.
Driscoll HT. *Filter Aids and Materials: Technology and Applications.* Park Ridge, NJ: Noyes Data Corp, 1977.
Hwang ST, Kammermeyer K. *Membranes in Separations,* vol 3. New York: Wiley Interscience, 1975.
Kolthoff IM, Elving PJ. *Treatise on Analytical Chemistry,* vol 5, pt 1. New York: Interscience, 1982.
Lachman L, et al. *The Theory and Practice of Industrial Pharmacy,* 3rd ed. Philadelphia: Lea & Febiger, 1986, chap 7.
Mink HP. *Application of a Multicomponent Membrane Transport Model to Reverse Osmosis Separation Processes.* ACS Symposium Series, no 281. Washington, DC: American Chemical Society, 1985.
Perry ES, Weissberger A. *Technique of Chemistry,* vol 12, 3rd ed. New York: Wiley Interscience, 1978.
Perry JH, et al. *Chemical Engineer's Handbook,* 6th ed. New York: McGraw-Hill, 1984.
Swarbrick J, Boylan J. *Encyclopedia of Pharmaceutical Technology.* New York: Dekker, 1990.
Townsend A. *Encyclopedia of Analytical Science.* New York: Academic, 1995.

Powders

Robert E O'Connor, PhD
Joseph B Schwartz, PhD
Linda A Felton, PhD

Powders are encountered in almost every aspect of pharmacy, both in industry and in practice. Drugs and other ingredients, when they occur in the solid state in the course of being processed into a dosage form, usually are in a more or less finely divided condition. Frequently, this is a powder whose state of subdivision is critical in determining its behavior both during processing and in the finished dosage form. Apart from their use in the manufacture of tablets, capsules, and suspensions, powders also occur as a pharmaceutical dosage form. Although the use of powders as a dosage form has declined, the properties and behavior of finely divided solid materials are of considerable importance in pharmacy. This chapter is intended to provide an introduction to the fundamentals of powder mechanics and the primary means of powder production and handling. The relationships of the principles of powder behavior to powders as dosage forms are discussed.

PRODUCTION METHODS

Molecular Aggregation

PRECIPITATION AND CRYSTALLIZATION

The precipitation and crystallization processes are fundamentally similar and depend on achieving three conditions in succession: a state of supersaturation (super cooling in the case of crystallization from a melt), formation of nuclei, and growth of crystals or amorphous particles.

Supersaturation can be achieved by evaporation of solvent from a solution, cooling of the solution if the solute has a positive heat of solution, production of additional solute as a result of a chemical reaction, or a change in the solvent medium by addition of various soluble secondary substances. In the absence of seed crystals, significant supersaturation is required to initiate the crystallization process through formation of nuclei. A nucleus is thought to consist of from 10 to a few hundred molecules having the spatial arrangement of the crystals that will be grown ultimately from them.

Such small particles are shown by the Kelvin equation to be more soluble than large crystals; therefore, they require supersaturation, relative to large crystals, for their formation and subsequent growth. It is a gross oversimplification to assume that, for a concentration gradient of a given value, the rate of crystallization is the negative of the rate of dissolution. The latter is generally somewhat greater.

Depending on the conditions of crystallization, it is possible to control or modify the nature of the crystals obtained. When polymorphs exist, careful temperature control and seeding with the desired crystal form are often necessary. The habit or shape of a given crystal form often highly depends on impurities in solution, pH, rate of stirring, rate of cooling, and the solvent. Very rapid rates of crystallization can result in impurities being included in the crystals by entrapment.

SPRAY-DRYING

Atomization of a solution of one or more solids via a nozzle, spinning disk, or other device, followed by evaporation of the solvent from the droplets is termed *spray-drying*. The nature of the powder that results is a function of several variables, including the initial solute concentration, size distribution of droplets produced, and rate of solvent removal. The weight of a given particle is determined by the volume of the droplet from which it was derived and by the solute concentration. The particles produced are aggregates of primary particles consisting of crystals and/or amorphous solids, depending on the rate and conditions of solvent removal. This approach to the powdered state provides the opportunity to incorporate multiple solid substances into individual particles at a fixed composition, independent of particle size, and avoiding difficulties that can arise in attempting to obtain a uniform mixture of several powdered ingredients by other procedures.

Particle-Size Reduction

Comminution in its broadest sense is the mechanical process of reducing the size of particles or aggregates. Thus, it embraces a wide variety of operations including cutting, chopping, crushing, grinding, milling, micronizing, and trituration, which depend primarily on the type of equipment employed. The selection of equipment in turn is determined by the characteristics of the material, the initial particle size and the degree of size reduction desired. For example, very large particles may require size reduction in stages simply because the equipment required to produce the final product will not accept the initial feed, as in crushing prior to grinding. In the case of vegetable and other fibrous material, size reduction generally must be, at least initially, accomplished by cutting or chopping.

Chemical substances used in pharmaceuticals, in contrast, generally need not be subjected to either crushing or cutting operations prior to reduction to the required particle size. However, these materials do differ considerably in melting point, brittleness, hardness, and moisture content, all of which affect the ease of particle-size reduction and dictate the choice of equipment. The heat generated in mechanical grinding, in particular, presents problems with materials that tend to liquefy or stick together and with the thermolabile products that may degrade unless the heat is dissipated by use of a flowing stream

of water or air. The desired particle size, shape, and size distribution also must be considered in the selection of grinding or milling equipment. For example, attrition mills tend to produce spheroidal, more free-flowing particles than do impact-type mills, which yield more irregular-shaped particles.

FRACTURE MECHANICS

Reduction of particle size through fracture requires application of mechanical stress to the material to be crushed or ground. Materials respond to stress by yielding, with subsequent generation of strain. Depending on the time course of strain as a function of applied stresses, materials can be classified according to their behavior over a continuous spectrum ranging from brittle to plastic. In the case of a totally brittle substance, complete rebound would occur on release of applied stress at stresses up to the yield point, where fracture would occur. In contrast, a totally plastic material would not rebound nor would it fracture.

The vast majority of pharmaceutical solids lie somewhere between these extremes and thus possess both elastic and viscous properties. Linear and, to a lesser extent, nonlinear viscoelastic theory has been developed well to account for quantitatively and explain the simultaneous elastic and viscous deformations produced in solids by applied stresses.

The energy expended by comminution ultimately appears as surface energy associated with newly created particle surfaces, internal free energy associated with lattice changes, and as heat. Most of the energy expressed as heat is consumed in the viscoelastic deformation of particles, friction, and in imparting kinetic energy to particles. Energy is exchanged among these modes and some is, of course, effective in producing fracture. It has been estimated that 1% or less of the total mechanical energy used is associated with newly created surface or with crystal lattice imperfections.

Although the grinding process has been described mathematically, the theory of grinding has not been developed to the point where the actual performance of the grinding equipment can be predicted quantitatively. However, three fundamental laws have been advanced:

Kick's Law—The work required to reduce the size of a given quantity of material is constant for the same reduction ratio regardless of the original size of the initial material.

Rittinger's Law—The work used for particulate size reduction is directly proportional to the new surface produced.

Bond's Law—The work used to reduce the particle size is proportional to the square root of the diameter of the particles produced.

In general, however, these laws have been useful only in providing trends and qualitative information on the grinding process. Usually laboratory testing is required to evaluate the performance of particular equipment. A work index, developed from Bond's Law, is a useful way of comparing the efficiency of milling operations.[1] A grindability index, which has been developed for a number of materials, also can be used to evaluate mill performance.[2]

A number of other factors also must be considered in equipment selection. Abrasion or mill wear is an important factor in the grinding of hard materials, particularly in high-speed, close-clearance equipment (eg, hammer mills). In some instances mill wear may be so extensive as to lead to highly contaminated products and excessive maintenance costs that make the milling process uneconomical. Hardness of the material, which often is related to abrasiveness, also must be considered. This usually is measured on the Moh's scale. Qualitatively, materials from 1 to 3 are considered as soft and from 8 to 10 as hard. Friability (ease of fracture) and fibrousness can be of equal importance in mill selection. Fibrous materials, such as plant products, require a cutting or chopping action and usually cannot be reduced in size effectively by pressure or impact techniques. A moisture content above about 5% will in most instances also create a problem and can lead to agglomeration or even liquefaction of the milled material. Hydrates often will release their water of hydration under the influence of a high-temperature milling process and thus may require cooling or low-speed processing.

METHODS AND EQUIPMENT

When a narrow particle-size distribution with a minimum of fines is desired, closed-circuit milling is advantageous. This technique combines the milling equipment with some type of classifier (see *Particle-Size Measurement and Classification*). In the simplest arrangement, a screen is used to make the separation, and the oversize particles are returned to the mill on a continuous basis while the particles of the desired size pass through the screen and out of the grinding chamber. Overmilling, with its subsequent production of fines, thereby is minimized. Equipment also has been designed to combine the sieving and milling steps into a single operation (see *Centrifugal-Impact Mills and Sieves*).

To avoid contamination or deterioration, the equipment used for pharmaceuticals should be fabricated of materials that are chemically and mechanically compatible with the substance being processed. The equipment should be easy to disassemble for cleaning to prevent cross-contamination. Dust-free operation, durability, simplified construction, and operation and suitable feed and outlet capacities are additional considerations in equipment selection.

Although there is no rigid classification of large-scale comminution equipment, it generally is divided into three broad categories based on feed and product size:

1. *Coarse crushers* (eg, jaw, gyratory, roll, and impact crushers).
2. *Intermediate grinders* (eg, rotary cutters, disk, hammer, roller, and chaser mills).
3. *Fine grinding mills* (eg, ball, rod, hammer, colloid, and fluid-energy mills; high-speed mechanical screen and centrifugal classifier).

Machines in the first category are employed ordinarily where the size of the feed material is relatively large, ranging from 1½ to 60 inches in diameter. These are used most frequently in the mineral crushing industry and will not be considered further. The machines in the second category are used for feed materials of relatively small size and provide products that fall between 20- and 200-mesh. Those in the third category produce particles, most of which will pass through a 200-mesh sieve, although often the particle size of the products from fine grinding mills is well into the micron range.

The comminution effect of any given operation can be described mathematically in terms of a matrix whose elements represent the probabilities of transformation of the various-size particles in the feed material to the particle sizes present in the output. The numerical values of the elements in the transition matrix can be determined experimentally and the matrix serves to characterize the mill. Matrices of this type are frequently a function of feed rate and feed particle-size distribution but are useful in predicting mill behavior. Multiplication of the appropriate comminution matrix with the feed-size distribution line-matrix yields the predicted output-size distribution.

INTERMEDIATE AND FINE GRINDING MILLS

The various types of comminuting equipment in this class generally employ one of three basic actions or, more commonly, a combination of these actions.

1. *Attrition.* This involves breaking down of the material by a rubbing action between two surfaces. The procedure is particularly applicable to the grinding of fibrous materials where a tearing action is required to reduce the fibers to powder.
2. *Rolling.* This uses a heavy rolling member to crush and pulverize the material. Theoretically, only a rolling-crushing type of action is involved, but in actual practice some slight attrition takes place between the face of the roller and the bed of the mill.
3. *Impact.* This involves the operation of hammers (or bars) at high speeds. These strike the lumps of material and throw them against each other or against the walls of the containing chamber. The impact causes large particles to split apart, the action continuing until small particles of required size are produced. In some instances high-velocity air or centrifugal force may be used to generate high-impact velocities.
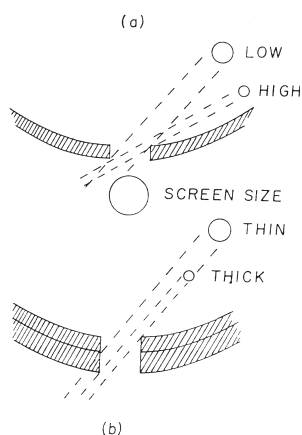
Figure 37-1. The influence of (**a**) mill speed and (**b**) screen thickness on particle size at a constant screen-opening size.[3]

Roller Mills—Roller mills in their basic form consist of two rollers revolving in the same direction at different rates of speed. This principle, which provides particle-size reduction mainly through compression (crushing) and shear, has been applied to the development of a wide variety of roller mills. Some use multiple smooth rollers or corrugated, ribbed, or saw-toothed rollers to provide a cutting action. Most allow adjustment of the gap between rollers to control the particle size of the product. The roller mill is quite versatile and can be used to crush a variety of materials.

An example of a pharmaceutical roller mill is the Crack-U-Lator, in which a series of ribbed rollers are adjusted to reduce sequentially the particle size of the product to produce the desired distribution. The design allows particles that are smaller than the gap between the rollers to pass to the next stage without unnecessary size reduction, thus reducing fines.

Hammer Mills—Hammer mills consist of a rotating shaft on which are mounted either rigid or swing hammers (beaters). This unit is enclosed with a chamber containing a grid or removable screen through which the material must pass. On the upper part is the feed hopper. As the material enters the chamber, the rapidly rotating hammers strike against it and break it into smaller fragments. These are swept downward against the screen where they undergo additional *hammering* action until they are reduced to a size small enough to pass through the openings and out. Oversize particles are hurled upward into the chamber where they also undergo further blows by the revolving hammers.

These mills operate at high speed and generally with controlled feed rate. Both impact and attrition provide the grinding action. Particle size is regulated by rotor speed, feed rate, type and number of hammers, clearance between hammers and chamber wall, and discharge openings. At a constant screen opening, the speed of the mill and the thickness of the screen will affect the particle size of the milled powder,[3] as shown in Figure 37-1. The higher the speed, the steeper the approach angle of the particle to the screen hole. Thus, for any screen size opening, the higher the blade speed, the smaller the particle obtained. Increasing the screen thickness will have a similar effect. In general flat-edged blades are most effective for pulverizing, while sharp-edged blades will act to chop or cut fibrous materials.

The FitzMill Comminutor (Fig 37-2) is an example of this type of mill. It can be used in either the hammer or knife-blade configuration and can be fitted with a wide range of screen sizes to fulfill a variety of milling specifications.

A wide range of particle sizes down to the micron size can be produced by these mills. The particle shape, however, is generally sharper and more irregular than that produced by compression methods. When very fine particles are desired, hammer mills can be operated in conjunction with an air classifier. Under such conditions a narrower particle-size distribution and lower grinding temperatures are obtained. Fine pulverizing of plastic material can be accomplished in these mills by embrittlement with liquid N_2 or CO_2 or by jacketing the grinding chamber.

Centrifugal-Impact Mills and Sieves—Centrifugal-impact mills and sieves are useful to minimize the production of fine particles, because their design combines sieving and milling into a single operation. The mill consists of a nonrotating bar or stator that is fixed within a rotating sieve basket. The particles that are smaller than the hole size of the sieve can pass through the mill without comminution; however, the particles or agglomerates larger than the hole size are directed by centrifugal force to impact with the stator. The sieve baskets also can be constructed to have a cutting edge that can aid in particle-size reduction without impact with the stator. The Quick Sieve (Fig 37-3), Turbo Sieve and CoMill are examples of this type of mill.

Cutter Mills—Cutter mills are useful in reducing the particle size of fibrous material and act by a combined cutting and shearing action. They consist of a horizontal rotor into which is set a series of knives or blades. This rotor turns within a housing, and into it are set stationary bed knives. The feed is from the top and a perforated plate or screen is set into the bottom of the housing through which the finished product is discharged. The particle size and shape is determined by the plate size, gap between rotor and bed knives, and size of the openings. A number of rotor styles are available to provide different particle shapes and sizes, though cutter mills are normally not designed to produce particles finer than 80- to 100-mesh.

Attrition Mills—Attrition mills make use of two stone or steel grinding plates, one or both of which revolve to provide grinding mainly through attrition. These mills are most suitable for friable or medium-hard, free-flowing material.

A double-runner attrition mill is an example of a mill that uses two rotating disks revolving in opposite directions. The particle-size reduction is controlled by varying the rotational speed of the disks, the space between the disks, and the size and number of ridges and indentations in the face of the disks. By appropriate combination with a classifier, particle sizes ranging from 10-mesh to 20 μm can be obtained by these attrition mills.

Figure 37-2. EZ-Clean FitzMill Comminutor (courtesy, Fitzpatrick).

Figure 37-3. Quick Sieve (courtesy, Glatt Air).

Figure 37-4. Single jar mill.

tating parts, and thus can be integrated readily into a particle classifying system or other ancillary equipment. Furthermore, there have been several studies that have demonstrated that the vibrating ball mill will grind at rates often as high as 20 to 30 times that of the conventional tumbling mill and offer a higher order of grinding rate and efficiency than other prevailing milling procedures.

Fluid-Energy Mills—Fluid-energy mills are used for pulverizing and classifying extremely small particles of many materials. The mills have no moving parts, grinding being achieved by subjecting the solid material to streams of high-velocity elastic fluids, usually air, steam, or an inert gas. The material to be pulverized is swept into violent turbulence by the sonic and supersonic velocity of the streams. The particles are accelerated to relatively high speeds; when they collide with each other, the impact causes violent fracture of the particles.

One type of fluid-energy mill is shown in Figure 37-5. The elastic grinding fluid is introduced through nozzles in the lower portion of the mill under pressures ranging from 25 to 300 psi. In this way, a rapidly circulating flow of gas is generated in the hollow, doughnut-shaped mill. A Venturi feeder introduces the coarse material into the mill and the particles enter into the jet stream of rapidly moving gas. The raw material is pulverized

Chaser Mills—Chaser mills are so called because two heavy granite stones, or chasers, mounted vertically like wheels and connected by a short horizontal shaft, are made to revolve or chase each other upon a granite base surrounded by a curb. Revolution of the chasers produces an upward current of air; this carries over the lighter particles, which fall outside the curb and subsequently are collected as a fine powder.

Pebble or Ball Mills—Pebble or ball mills, sometimes called *pot mills* or *jar mills,* are operated on the principle of attrition and impact. The grinding is effected by placing the substance in jars or cylindrical vessels that are lined with porcelain or a similar hard substance and containing *pebbles* or *balls* of flint, porcelain, steel, or stainless steel. These cylindrical vessels revolve horizontally on their long axis and the tumbling of the pebbles or balls over one another and against the sides of the cylinder produces pulverization with a minimum loss of material. Ball-milling is a relatively slow process and generally requires many hours to produce material of suitable fineness. To keep the grinding time within reasonable limits, coarse material (>10-mesh) should be preground before introduction into a ball mill. Figure 37-4 shows a sectional view of a single jar mill. *Rod mills* are a modification in which rods about 3 inches shorter than the length of the mill are used in place of balls. This results in a lower production of fines and a somewhat more granular product.

Vibrating Ball Mills—Vibrating ball mills, which also combine attrition and impact, consist of a mill shell containing a charge of balls similar to rotating ball mills. However, in this case the shell is vibrated at some suitable frequency, rather than rotated. These mills offer the advantage of being free of ro-
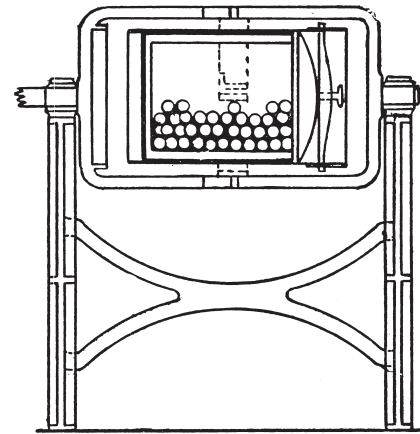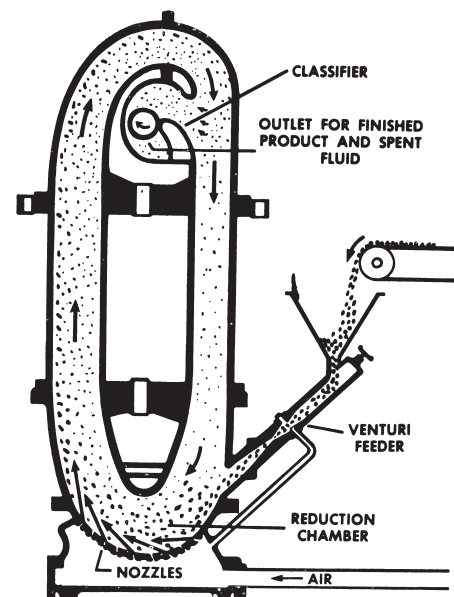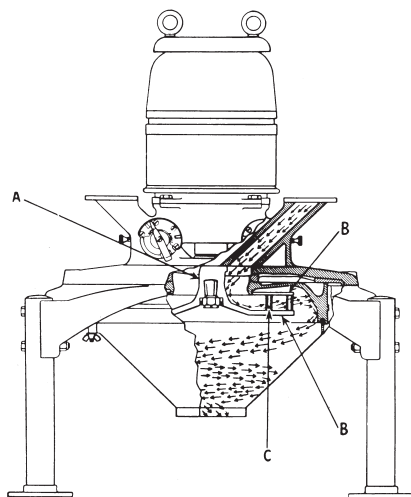

Figure 37-5. The Jet-O-Mizer fluid energy mill (courtesy, Fluid Energy).

Figure 37-6. CentriMil, a centrifugal-impact mill, available in models ranging from 2 to 250 hp. **A.** Spinning roto. **B.** Rotor hub disks. **C.** Impacters (courtesy, Entoleter).

Table 37-1. Definition of Statistical Diameters[a]

TYPE OF MEAN DIAMETER	STATISTICAL DEFINITION	DESCRIPTION
Arithmetic	$\sum nd/\sum n$	Mean diameter weighted by number
Diameter moment	$\sum nd^2/\sum nd$	Mean diameter weighted by particle diameter
Surface moment	$\sum nd^3/\sum nd^2$	Mean diameter weighted by particle surface
Volume moment	$\sum nd^4/\sum nd^3$	Mean diameter weighted by particle volume
Surface	$(\sum nd^2/\sum n)^{1/2}$	Root mean square
Volume	$(\sum nd^3/\sum n)$	

[a] When grouped data are used, *n* is the number of particles in a size interval characterized by a diameter, *d*.

quickly by mutual impact in the reduction chamber. As the fine particles form, they are carried upward in the track. Particles are ground simultaneously and classified in this process. The smaller particles are entrapped by the drag of gas leaving the mill and are carried out to a collecting chamber or bag. Centrifugal force at the top of the chamber stratifies the larger, heavy particles and their greater momentum carries them downward and back to the grinding chamber.

A major advantage of the fluid-energy mill is that the cooling effect of the grinding fluid as it expands in the grinding chamber more than compensates for the moderate heat generated during the grinding process. Another advantage is the rather narrow range of particle sizes produced. When precise control of particle size is an important factor, the fluid-energy mill produces very narrow ranges of particles with minimum effort.

One major disadvantage is the necessity of controlling the feeding of the coarse, raw material into the jet stream. Often, the feeding device becomes clogged by a clump of material, and special feeding devices must be built to produce a uniform rate of feed.

Centrifugal-Impact Pulverizers—Centrifugal-impact pulverizers also have been found to be effective for the reduction of the particle size of a wide variety of materials ranging from very soft, organic chemicals to hard, abrasive minerals. In addition, this type of mill is suited well for the size reduction of heat-sensitive substances. Basically, in these pulverizers, the material is fed into the center of a spinning rotor that applies a high centrifugal force to the particles. The material, thus accelerated, moves toward the impactor set at the periphery of the rotor. On striking these impactors the material is hurled against the outer casing where final reduction is achieved. Processed material is removed from the bottom of the conical discharge hopper (Fig 37-6). Particle-size reduction in the range of 10- to 325-mesh can be obtained with this type of mill with a minimum of fines.

PARTICLE-SIZE MEASUREMENT AND CLASSIFICATION

Size and Distribution

STATISTICAL PARAMETERS

Monodisperse systems of particles of regular shape, such as perfect cubes or spheres, can be described completely by a single parameter: the length of a side or diameter. However, when either nonuniform size distributions or anisometric shapes exist, any single parameter is incapable of fully defining the powder. Measurements must be made over the total range of sizes present. Statistical diameters, for example, are useful measures of central size tendency and are computed from some measured property that is a function of size and related to a linear dimension. For irregular particles the assigned size will depend strongly on the method of measurement.

Once a method of assignment of numerical value for the diameter, surface area or other parameter has been established, the average value computed depends on the weighting given the various sizes. Mean particle diameter is the most important single statistical parameter because, if the proper diameter is chosen, the various other parameters of interest such as specific surface area, number, mean particle weight often may be calculated. Thus, the choice of the mean diameter to be measured or calculated is based on its intended use. For example, specific surface area, which may control drug dissolution, frequently can be related to the root-mean square diameter. Depending on the method of measurement, various diameters are obtained; these will be discussed later. The particle diameters most commonly used are listed in Table 37-1.

SIZE DISTRIBUTIONS

As has been pointed out, size distributions are often complex and no single particle-size parameter is sufficient to characterize or permit prediction of the many bulk properties of pharmaceutical interest, such as flow characteristics, packing densities, compressibility, or segregation tendencies. Thus, descriptions beyond the central tendency provided by the various mean diameters are needed. These generally take the form of equations or charts that describe in detail the distribution of particle size. In measuring particle size it is important first to select the parameter that is related to the ultimate use of the product, and then select the method that will measure this parameter.

Certainly, more useful information would be gained if the particle size of a powder used in a suspension were determined by sedimentation rather than by microscopy, or if the total surface area of the particles were the critical factor (as in use as an adsorbant) by the more useful method of permeability or gas adsorption.

Particles can be classified by determining the number of particles in successive size ranges. The distribution can be represented by a bar graph or histogram (Fig 37-7), where the widths of the bars represent the size range and the heights represent the frequency of occurrence in each range. A smooth curve drawn through the midpoints of the tops of the bars in this case results in a normal probability size-distribution curve. A line drawn through the center of the curve to the abscissa divides the area into two equal parts and represents the mean value. Because a number of other symmetrical distributions could have this same midpoint, a term to describe the scatter
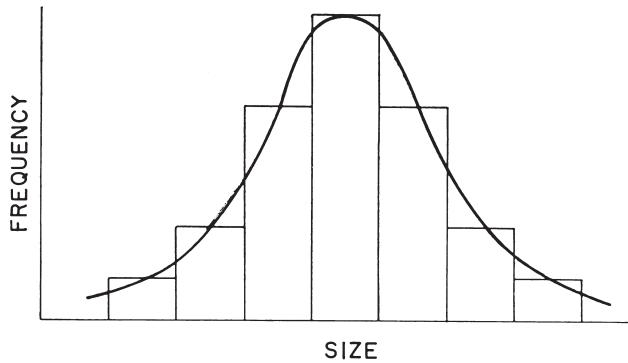
Figure 37-7. Symmetrical particle-size distribution curve.

about the mean value is needed. Standard deviation (the root-mean square deviation about the mean) serves to define the spread of the curve on either side of the midpoint.

Most particulate material cannot, however, be described by a normal distribution curve. The resultant curves are usually skewed as shown in Figure 37-8, making mathematical analysis complex. In a skewed size distribution, the mean value is affected by very large or very small values. In these cases, the median (ie, the central value of a series of observations) is a more useful average. In a symmetrical distribution the mean and the median values are the same. Most asymmetrical size distribution curves relating to powders can be converted into symmetrical curves by using the logarithm of the size—the log normal distribution curve. The symmetrical shape of the latter curve allows for simplified mathematical analysis.

Cumulative plots are also useful for particle-size distribution analysis. Here, the cumulative percent of the particles that are finer (or larger) than a given size is plotted against the size. By use of logarithmic-probability paper, the median size (geometric mean) and standard deviation (geometric standard deviation) can be obtained readily by graphical solution. The median is the 50% size and the standard deviation is the slope of the line and equal to the ratio 50% size/15.87% size (Fig 37-9).

Size Measurement

Frequently, particle-size measurements are made in conjunction with separation of the powder into fractions on the basis of size. Methods that lead primarily to size distribution analysis only are discussed first, followed by methods in which classification by size is a central feature.

The basic processes employed for measurement, classification, or fractionation of fine solid particles involve direct and indirect techniques. Direct methods measure the actual di-

mensions of the particle by use of a calibration scale, as in microscopy and sieving. Indirect measurements make use of some characteristic of the particle that can be related to particle size, such as sedimentation rates, permeability, and optical properties.

MICROSCOPY

Microscopic techniques have been classified as one of the most accurate of *direct* methods. Here, particles are sized directly and individually, rather than being grouped statistically by some other means of classification. The linear measurement of particles is made by comparison with a calibrated scale usually incorporated into the microscope. For spherical particles the size is defined by the measurement of the diameter. However, for other-shaped particles some alternative single size designation is generally used, such as the diameter of a sphere with the same projected area as the nonspheroidal particle being measured. Other characteristic diameters based on various aspects of the projected particle outline as seen through the microscope also have been reported in the literature to describe non-spheroidal particles.

The method is rather tedious and other limitations are found in the techniques required for preparation of the slides and in the maximum resolution that sets the lower limits of particle size measurement using visible light. White light can resolve particles within the range of 0.2 to 100 μm. This lower limit can be decreased to about 0.1 μm by the use of ultraviolet

Figure 37-9. Log probability plot of particle size versus cumulative weight % frequency oversize.

Figure 37-8. Skewed particle-size distribution curve.

light and to about 0.01 μm by the use of the ultramicroscope. The electron microscope finds its greatest usefulness in particle-size measurements in the range of 0.001 to 0.2 μm.
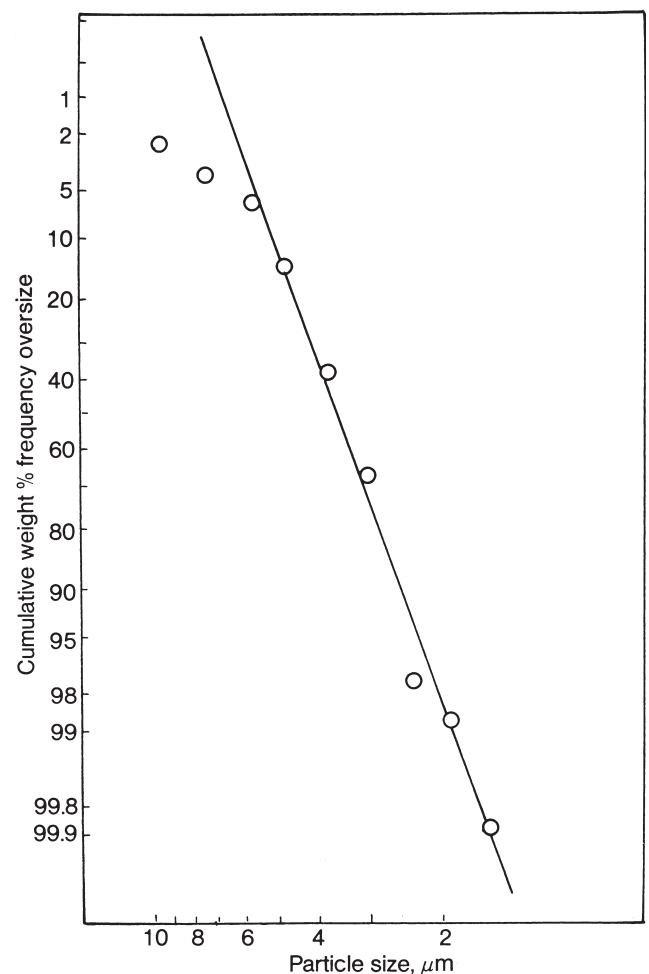
Although microscopic methods for particle size determination are time consuming, tedious, and generally require more skill than some of the other techniques, they offer a number of advantages. They supply information about particle shape and thickness that cannot be obtained by other methods and, in addition, supply a permanent record through use of photomicrographs.

A variety of semiautomated procedures have been developed to reduce the fatigue and tedium associated with manual counting of particles. These are represented by instruments such as the Imanco Quantimet 720 and the πMC System (*Millipore*), which scan the powder image in a manner similar to a TV scanner. The signal obtained is analyzed by a pulse-height analyzer and expressed as a particle-size distribution.

ADSORPTION OF GASES

Adsorption of a solute from solution or of a gas at low temperatures onto powdered material serves as a measure of the particle surface area, generally reported as specific surface (area/unit mass). Common adsorption techniques use the adsorption of nitrogen and krypton at low temperatures. The volume of the gas adsorbed by a powdered sample is determined as a function of gas pressure, and an appropriate plot is prepared. The point at which a monomolecular layer of adsorbate occurs is estimated from the discontinuity that shows in the curve. The specific surface area then can be calculated from knowledge of the volume of gas required to achieve this monolayer, and the area/molecule occupied by the gas, its molecular weight and density. Frequently, more complex expressions such as the Brunauer, Emmett, and Teller (BET) equation must be used to describe the surface adsorption of some materials and determine the volume of gas required to produce an adsorbed monolayer. The surface properties of a number of pharmaceuticals have been investigated by this technique.

PERMEABILITY

When a gas or liquid is allowed to flow through a powdered material, the resistance to this flow is a function of such factors as specific surface of the powder, area of the bed, pore space, pressure drop across the bed, and viscosity of the fluid. This resistance can be described and the specific surface calculated by the Kozeny–Carmen equation, which relates these factors. This method, although it does not provide a size distribution analysis, offers a rapid and convenient means of size estimation that is useful for some industrial operations.

Instruments that measure the rate of flow of a gas through a powder bed under controlled pressure differential are available commercially. The Sub-Sieve Sizer (*Fisher*) permits the reading of average particle size directly. The Blaine Permeameter (*Precision Scientific*) uses the principle of filling the void spaces in a powder with mercury and then weighing it. The void fraction is calculated from the known density of mercury at different temperatures.

The calculations involved in permeability techniques are often complicated and yield only an average size of particles. In measuring particles in the subsieve ranges, rather large deviations may be encountered. With larger mesh sizes, some good agreement is found between the results obtained by techniques employing permeability and microscopy, particularly if the powders are made up of spherical or near-spherical particles.

IMPACTION AND INERTIAL TECHNIQUES

The laws that govern the trajectories of particles in fluid streams are used in several methods of particle-size measure-

ment. Impaction devices are based on the dynamics of deposition of fine particles in a moving air stream when directed past obstacles of defined geometric form, or when forced from a jet device onto a plane surface.

The *cascade impactor,* described by Pilcher and co-workers,[4] forces particle laden air at a very high speed and fixed rate through a series of jets (each smaller than the preceding one) onto glass slides; impaction takes place in a series of stages. The velocities of the air stream and the particles suspended in it are increased as they advance through the impactor. As a result, the particles are classified by impaction on the different slides, with the larger particles on the top slides and the smaller ones on the downstream slides. Figure 37-10 illustrates the principle of the cascade impactor. The exact size of impacted particles on each slide subsequently must be determined. Size analyses may be obtained directly by theoretical treatment or prior calibration of the instrument.

Tillotson[5] described an instrument based on inertial principles similar to those of the cascade impactor. This instrument may be adapted for automatic readout of size distribution by means of light-scattering techniques and electronic counters. The method is claimed to provide complete particle-size distribution data in a few minutes.

AUTOMATIC PARTICLE SIZE COUNTERS

The principles of electronic and light sensing and light scattering techniques have been used to develop automated particle size counters that indirectly measure particle size.

Electrozone Sensing—The *Coulter Counter* determines the particle volume distribution of materials suspended in an electrolyte-containing solution. This instrument utilizes an electrical sensing zone and measures electrical pulses caused by the passage of particles through the zones. The instrument must be calibrated with monodispersed particles of known diameter. A table of size ranges of several methods
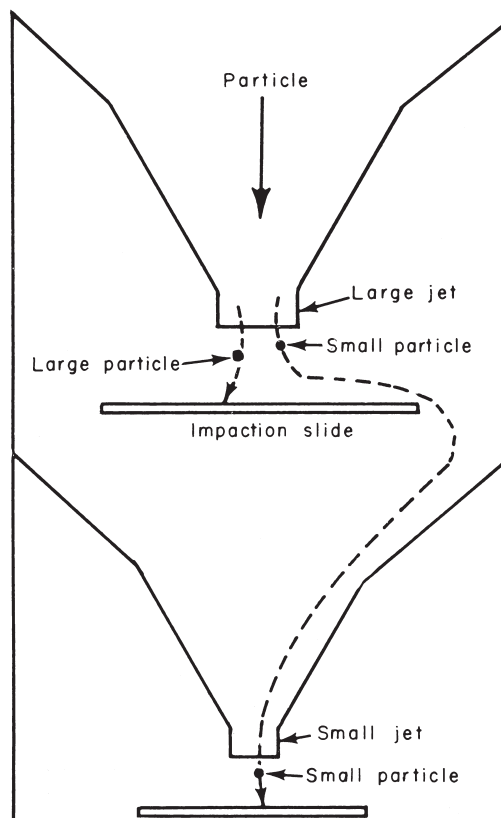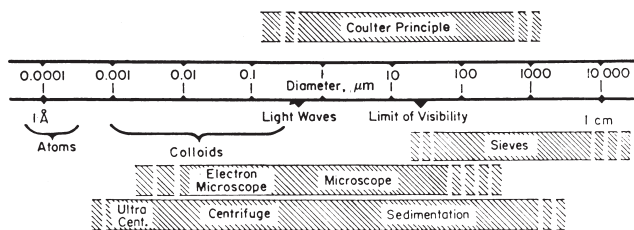

Figure 37-10. The principle of the cascade impactor.

Figure 37-11. Size range of Coulter method compared with coverage of sieve, sedimentation, and microscopic methods, and overlap of electron microscope and centrifuge ranges (courtesy, Coulter).

compared with the Coulter principle is shown in Figure 37-11. Detection is limited by thermal and electrical noise and the ability to discriminate true signal pulses from background.

Photozone Sensing—The *HIAC Counter* measures the size distribution of particles suspended in either liquids or gases. The standard models will measure sizes from 2 to 2500 μm at pressures up to 3000 psi. Basically, in this instrument the particles pass a window, one by one. As each particle passes, depending on its size it interrupts some portion of a light beam. This causes an instantaneous reduction in the voltage from a photodetector that is proportional to the size of the particle. Several counting circuits with preset thresholds tally the particles by size.

Laser Diffraction—Laser diffraction or low angle laser light scattering has become one of the preferred methods for particle size characterization. The instrument consists of a laser light source (generally a He-Ne gas laser), a suitable detector such as a silicon photodiode, and a means of passing the sample through the laser beam. An ultrasonic probe may be used to improve particle dispersion. In this technique, particles are dispersed in a liquid or gaseous medium. The diffraction angle is inversely proportional to particle size. The instrument does not require calibration against a standard and dry powders may be measured directly by using pressure to pass the sample through the instrument. In addition, the method is nondestructive and samples can be recovered after testing. The latest instruments utilize the Mie theory of particle interaction with light and allow for accurate measurements over a large size range (typically 0.1 to 3000 μm)[6].

Size Classification

SIEVING

Sieving is one of the simplest and probably most frequently used methods for determining particle-size distribution. The technique basically involves size classification followed by the determination of the weight of each fraction.

In this technique, particles of a powder mass are placed on a screen made of uniform apertures. By the application of some type of motion to the screen, the particles smaller than the apertures are made to pass through. The sieve motion generally is either (1) horizontal, which tends to loosen the packing of the particles in contact with the screen surface, permitting the entrapped subsieve particles to pass through, or (2) vertical, which serves to agitate and mix the particles as well as to bring more of the subsieve particles to the screen surface.

One major difficulty associated with this method is the production of screens with uniform apertures, particularly in the very fine mesh sizes. As a result the practical lower limit for woven-wire mesh screens is about 43 μm (325-mesh). However, with the introduction of electroformed screens, sieves capable of analyzing particles in the 5-μm range are now available. In addition, "blinding" of the openings by oversized or irregular particles and inefficient presentation of the particles to the screen surface are problems associated with this technique. The use of horizontal and vertical screening motions, air jets, sudden periodic reversal of the sieve motion, and continuous cycling all have been used in an attempt to eliminate these problems.

For continuous operations, the screens are attached to mechanical or electromagnetic devices that supply the energy required to shake the particles through the openings in the screen and also to prevent accumulation of fines within the openings, as this tends to clog them and slow down the operation. The use

of an electromagnetic instead of mechanical drive provides a more gentle sieving action with a resultant decrease in sieve wear, blinding, and machine noise. Sieves may be used either in a sequence of sizes through which the material must pass or singly in the required size.

This apparatus is useful in obtaining size-analysis data under controlled conditions. The sample is placed in the top of the nest of standard sieves arranged in a descending order. The length of time and force of vibration to which the sample is subjected may be preset by variable time and voltage controls. The controlled vibration causes the powder particles to pass through the sieves, each fraction coming to rest in the sieve through which it cannot pass. For the purpose of analysis, the weight of each fraction is determined and the percentage calculated.

The Sonic Sifter (*Allen-Bradley* and *ATM*) is a laboratory sifter that uses sonic oscillation to classify particles. A mechanical pulse action is used to reduce blinding and agglomeration in the subsieve sizes. This combination of sonic and mechanical agitation permits dry sifting down to 5 μm. US Standard Sieves are available for this unit from 31/2- to 400-mesh and in precision electroformed mesh sizes from 150 to 5 μm.

Industrial-size mechanical sieves are varied in design and capacity, and include the gyratory, circular rotatory, vibrating, shaking, and revolving sifters. In gyratory sifters, the motion is in a single horizontal plane, but may vary from circular to reciprocal from the feed to the discharge end. The circular sifter also confines the screen motion to a horizontal plane, but in this case the total motion applied to the sieve is circular. The material enters the top of a gyratory sifter and spreads over the first sieve. Some of the finer particles drop through and are discharged into the *throughs* channel. The remaining powder moves to the next sieve in order, the process is repeated until complete separation is accomplished (Fig 37-12).
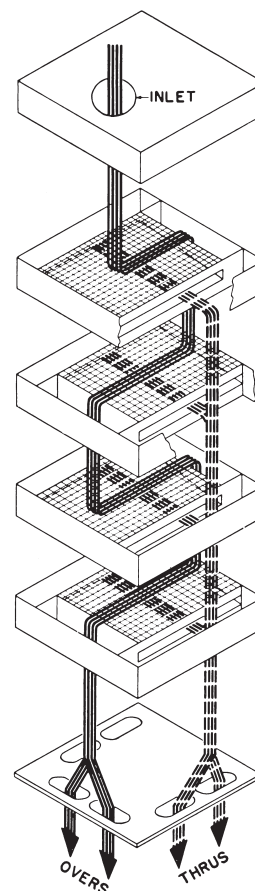
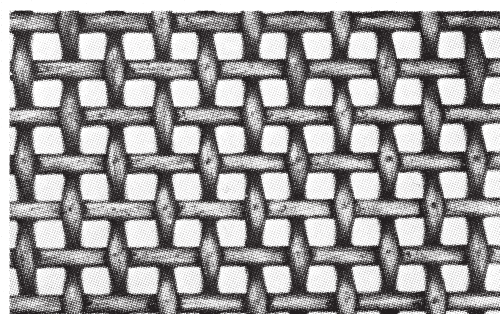**Figure 37-12.** Gyratory sifter (courtesy, Sprout Waldron).

Figure 37-13. Plain weave screen.

In centrifugal screening, the material is pushed through a spinning vertical wire cloth cylinder. Sharp cuts in particle size can be obtained with this type of equipment. Downward air flow, instead of shaking and tapping, has been used to move the particles through the screen openings; alternating with a reverse air flow serves to prevent *blinding,* particularly with fine-mesh sieves.

WET SCREENING

The addition of water sometimes is employed to dissolve any unwanted binders, remove fines or surface contamination, and to reduce surface forces—particularly in micromesh sieves—that oppose the flow of particles through the sieve. Particles that tend to agglomerate or react with oxygen or moisture and thus cannot be dry-sieved often can be handled by wet-sieving. Particles in the 6- to 150-µm range have been classified with good precision using electroformed sieves. Some hydrophobic substances that resist wetting by water may be wet screened by the use of organic liquids such as petroleum ether, acetone, or alcohol. Wet screening may be accomplished by spraying both the screen surface and the material as it is fed onto the screen or by feeding a slurry of material directly onto the screen.

SCREENING SURFACES

A number of factors must be considered in selecting screening surfaces. Primary consideration is given to the size and shape of the aperture opening, the selection of which is determined by the particle size that is to be separated. Screens commonly used in pharmaceutical processing include *woven wire screens, bolting cloth, closely spaced bars,* and *punched plates.* Punched plates are used for coarse sizing; their holes may be round, oval, square, or rectangular. The plates must be sturdy and withstand rough service. Sizes in common use range upward from 1/4 inch.

Most screening, however, is accomplished with woven-wire screens ranging in size from those with 400 openings/inch to screens with 4-inch square openings or larger. There are numerous types of woven wire screens, including plain, twilled, and braided weave. An example of the plain and twilled weave is shown in Figures 37-13 and 37-14.

In the US, the two common standards are the *Tyler Standard* and *US Standard* sieves. In both these series the sieve number refers to the number of openings per linear inch. For most purposes, screens from the two series are interchangeable, though in a few instances the number designations are different. Because these numbers do not define the size of the openings, the Bureau of Standards has established specifications for *Standard Sieves,* as given in Table 37-2. These specifications also establish tolerances for the evenness of weaving, as irregularities from careless weaving might permit much larger particles to pass the sieve than would be indicated. The standard sieves used for pharmaceutical testing are of wire cloth.

SEDIMENTATION

The sedimentation method employs the settling of particles in a liquid of a relatively low density, under the influence of a gravitational or centrifugal field. In free settling (ie, no particle-particle interference), the particles are supported by hydraulic forces and their fall can be described by Stokes' law. However, in most real situations, particle–particle interference, nonuniformity, and turbulence are all present, resulting in more complex settling patterns. The Andreason pipet, which is based on sampling near the bottom of a glass sedimentation chamber, is perhaps the best known of the early instruments. With centrifugation, entrainment of particles in the currents produced by other particles also may interfere with fractionation.

Gravitational settling chambers often are used for large-scale separation of relatively coarse particles in the range of 100 µm. Centrifugal devices are useful for the separation of much smaller particles (5–10 µm).

Sedimentation balances are available that provide a means of directly weighing particles at selected time intervals as they fall in a liquid system. For continuous observations, automatic recording balances also are available. A commercially available instrument called a *Micromerograph* uses the principle of sedimentation in an air column. This instrument and others related
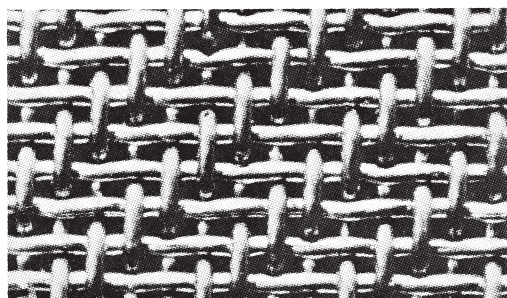
Figure 37-14. Twilled weave screen.

Table 37-2. Nominal Dimensions of Standard Sieves

| NO | SIEVE OPENING | | PERMISSIBLE VARIATION IN AVERAGE OPENING, % | PERMISSIBLE VARIATION IN MAXIMUM OPENING, % | WIRE DIAMETER, mm |
	mm	µm			
2	9.52	9520	±3	+5	2.11 to 2.59
4	4.76	4760	±3	+10	1.14 to 1.68
8	2.38	2380	±3	+10	0.74 to 1.10
10	2.00	2000	±3	+10	0.68 to 1.00
20	0.84	840	±5	+15	0.38 to 0.55
30	0.59	590	±5	+15	0.29 to 0.42
40	0.42	420	±5	+25	0.23 to 0.33
50	0.297	297	±5	+25	0.170 to 0.253
60	0.250	250	±5	+25	0.149 to 0.220
70	0.210	210	±5	+25	0.130 to 0.187
80	0.177	177	±6	+40	0.114 to 0.154
100	0.149	149	±6	+40	0.096 to 0.125
120	0.125	125	±6	+40	0.079 to 0.103
200	0.074	74	±7	+60	0.045 to 0.061

to it in principle offer more rapid determinations than those that use a liquid medium. There are, however, serious uncertainties in the method that must be taken into consideration. Deviations from Stokes' law and impaction of particles against the inner wall of the settling chamber are sources of possible error.

The Carey and Stairmand *photosedimentometer* photographs the tracks of particles as they fall in a dispersion medium. The size determination is derived from the length of the photographic track, which is an indication of the distance traveled by the particles, and the time of exposure of the photograph.

ELUTRIATION

In elutriation, the particles are suspended in a moving fluid, generally water or air. In vertical elutriation at any particular velocity of the fluid, particles of a given size will move upwards with the fluid, while larger particles will settle out under the influence of gravity. In horizontal elutriation a stream of suspended particles is passed over a settling chamber. Particles that leave the stream are collected in the bottom of the chamber. Normally, for all elutriation techniques, both undersize and oversize particles appear in each fraction and recycling is required if a clean cut is desired. By varying the fluid velocities stepwise, the sample may be separated into fractions. The amount in each fraction then can be determined and the size limits calculated by the use of the Stokes' equation or measured directly by microscopy. Air elutriation usually will give a sharper fractionation in a shorter time than will water elutriation.

Centrifugal elutriation is basically the same process, except in this case the fluid stream is caused to spin so as to impart a high centrifugal force to the suspended particles. The particles that are too large to follow the direction of flow separate out on the walls or bottom of the elutriator or cyclone. The finer particles escape with the discharge stream. Separation down to about 0.5 μm can be achieved with some centrifugal classifiers.

The DorrClone (*Dorr-Oliver*) (Fig 37-15) is an example of a centrifugal-type classifier. The feed enters tangentially into the upper section. Centrifugal forces in the vortex throw the coarser particles to the wall where they collect and then drop down and out of the unit. The fine particles move to the inner spiral of the vortex and are displaced upward and finally out of the top of the unit.

Inertial elutriators, which use an abrupt change in direction of the fluid stream to produce separation, are effective down to about 200-mesh. However, as with other elutriators, a clean cut usually cannot be obtained without recycling.

Felvation is a unique process that combines elutriation and sieving along with a varying fluid flow rate and a turbulent fluidized bed to achieve particle separation. The particles are fluidized within the felvation column. With a gradual increase of the fluid flow rate, the very fine particles are brought up to and then through a sieve surface set into the upper section of the column. These fines are filtered subsequently out of the fluid stream. A further increase in the fluid flow rate causes larger and larger particles to move through the sieve. The final stage is reached when particles just larger than the sieve aperture are elutriated up to the sieve.

Because of the way in which the particles are presented to the sieve, very little blinding of the openings occur. Furthermore, because the sieve need only serve as a go/no-go gauge and not as a supporting surface for the powder, a relatively small sieve surface is required. Thus, the more-uniform but more-expensive electroform sieves, even down to a 10-μm size, can be used in this process.

MISCELLANEOUS METHODS

Numerous other methods have been applied to particle size determination, including x-ray and electron diffraction, ultrasound, flotation, and electrostatic, magnetic, and dielectrophoretic methods. Newer techniques include photon
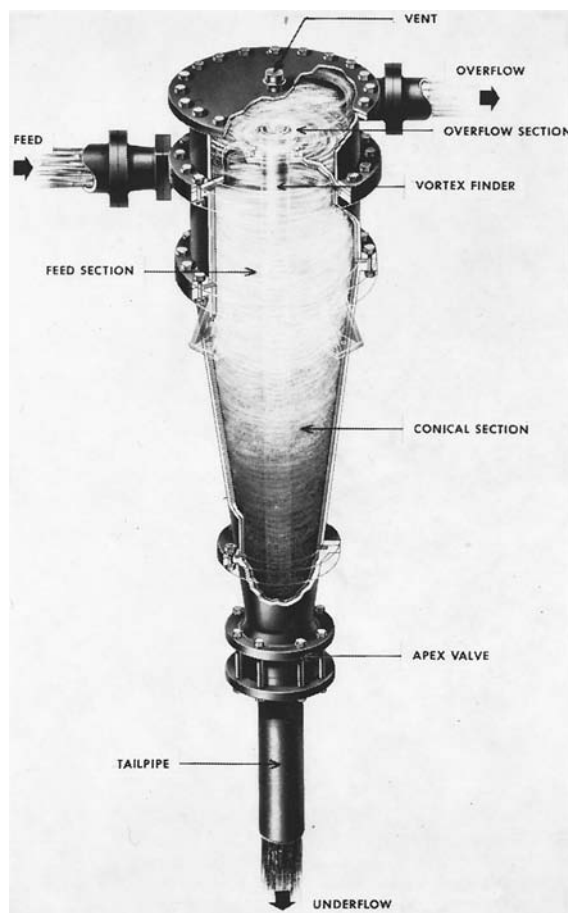
Figure 37-15. DorrClone, a hydrocentrifugal classifier (courtesy, Dorr-Oliver).

correlation spectroscopy, polarization intensity differential scattering, and fourier-transform infrared spectroscopy with diffuse reflectance. These techniques either are used principally as research tools or are industrial-scale methods of use outside the pharmaceutical industry. Detailed descriptions of their principles of operation and their applications can be found in the *Bibliography*.

SOLIDS HANDLING

Packing and Bulk Properties

BULK DENSITY; ANGLES OF REPOSE

Systems of particulate solids are the most complex physical systems encountered in pharmacy. No two particles in a powder are identical and the nature of momentum and energy exchange between particles defies description except in the most idealized and approximate terms. Bulk properties of powders are determined in part by the chemical and physical properties of their component solids and in part by the manner in which the various components interact. These interactions in turn frequently depend on the past history of the powder bed as well as on the ambient conditions.

The static properties of a particulate bed depend on particle–particle interactions and, in particular, on the way in which applied stresses are distributed through the bed. The number of contacts between particles and, hence, the average number of interparticulate contact points per particle increases as bed-packing increases. Packing may be expressed in terms of porosity,
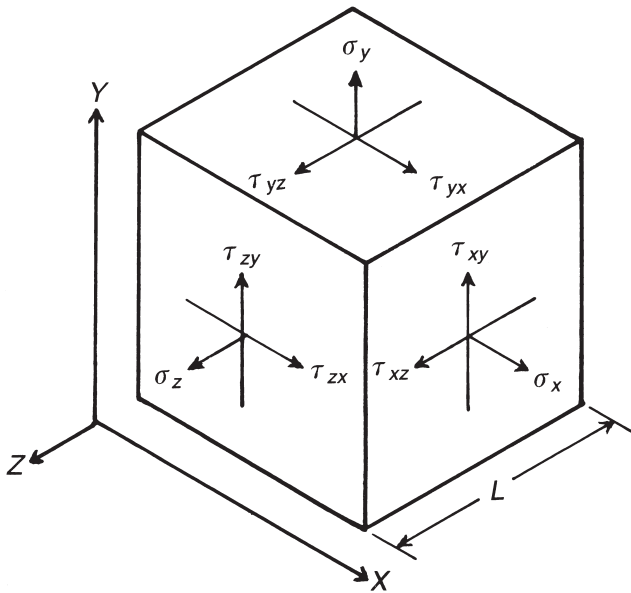
Figure 37-16.

percent voids, or fraction of solids by volume. Packings for regular arrangements of uniform spheres can be calculated and range in fractional solids from 0.53 for cubic to 0.74 for tetrahedral lattices. Powders composed of irregular-shaped particles in a distribution of sizes can pack to fractional densities approaching unity.

The manner in which stresses are transmitted through a bed and the bed's response to applied stress are reflected in the various angles of friction and repose. The most commonly used of these is the angle of repose, which may be determined experimentally by a number of methods, with slightly differing results. The typical method is to pour the powder in a conical heap on a level, flat surface and measure the included angle with the horizontal. Angles of repose range from 23° for smooth uniform glass beads to 64° for granular limestone. Cohesive materials frequently behave in an anomalous manner, yielding values in excess of 90°.

The angle of internal friction is a measure of internal stress distributions and is the angle at which an applied stress diverges as it passes through the bed. This angle together with the angle of slide are useful parameters in the design of storage/discharge bins. The latter angle is defined as the least slope at which a powder will slide down an inclined plane surface. Various other angles are in lesser use and will not be discussed here.

STATICS

Powders at rest experience stresses that vary with location throughout their volume and arise from pressures exerted by the container as well as from the weight of the bed above. Each point within the bed experiences both normal and shear stresses in general. Normal stresses may be either tensile or compressive. The powder bed will remain motionless and no flow will occur unless the normal and/or the shear strength is exceeded at some point within the bed. In general, the yield strengths, both normal and shear, are functions of the normal and shear stresses at the point of interest and depend upon the orientation of the axes of reference and the nature of the powder itself. It is apparent that to understand powder flow it is necessary to understand the conditions under which bed failure occurs and powder flow is initiated and sustained.

Consider the stresses that are applied to the faces of a small cube that is centered about a point chosen at random within a powder bed. Normal stresses are designated σ_i, where the subscript indicates the axis normal to the face and shear stresses are designated τ_{ij}, where the first subscript indicates the face and the second indicates the direction of the applied force. If the cube has an edge length, L, which is not infinitesimal, and if a stress gradient exists within the region, the corresponding stresses on opposite faces of the cube will not be equal. However, if the cube is made progressively smaller, and as L approaches zero, the stress values will converge to those at the point of interest. These forces are illustrated in Figure 37-16. It can be seen from this diagram that the state of stress at a point can be described by nine stress components.

If the system is in static equilibrium, and is not being accelerated translationally or rotationally, the forces that otherwise would result in movement must be in balance and have the effect of canceling each other. For example, τ_{xy} must equal τ_{yx} if rotation about the z-axis is not to occur. In a similar manner, shear and normal stresses, which would lead to translational movement along any of the three axes, also must balance.

Because the directions of the mutually perpendicular axes in Figure 37-16 were chosen arbitrarily, any other orientation of the cube corresponding to another set of axes also must result in a balance of forces. However, the distribution of stress among normal and shear components will depend on the particular axes selected. Thus, the stress condition of a powder can be analyzed in terms of the dependence of the normal and shear stresses on the direction chosen for the reference axes. This can be done by a method of analysis devised by Mohr, and can be visualized using a Mohr circle diagram, which permits stresses at any given point within a powder bed to be graphically resolved into normal, σ, and shear, τ, stresses for any arbitrary choice of axes.

For simplicity, assume that stress in the z-direction is not a function of z and that stress gradients exist in the x and y directions only. Stresses then can be analyzed in the xy plane without reference to the z-axis. Figure 37-17 shows the relationship between stresses relative to two xy coordinate systems at an angle θ to each other. If the condition of stress in the powder remains constant and only the angle θ between the two sets of reference axes is allowed to change, the resolution of stress into normal and shear components will be different for each set of axes and will depend on θ. By means of
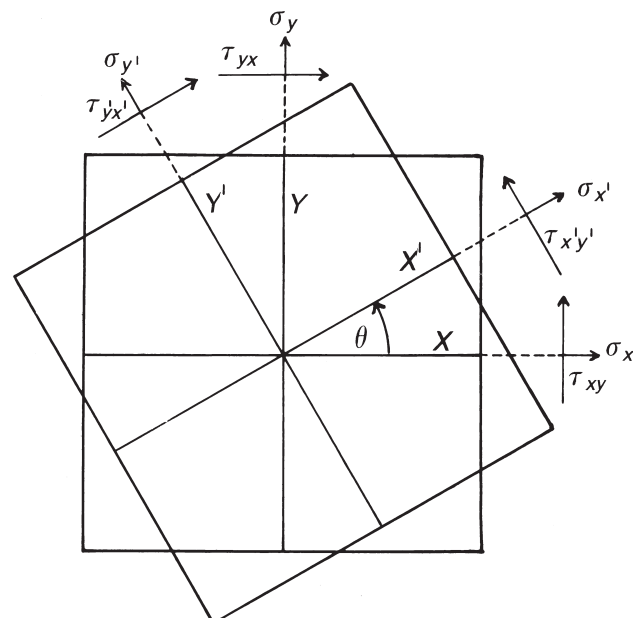

Figure 37-17.

trigonometry, the relationships between these two sets of stresses is shown to be

$$\sigma_{x'} = \frac{\sigma_x + \sigma_y}{2} + \frac{\sigma_x - \sigma_y}{2}\cos2\theta + \tau_{xy}\sin2\theta$$

$$\sigma_{y'} = \frac{\sigma_x + \sigma_y}{2} - \frac{\sigma_x - \sigma_y}{2}\cos2\theta - \tau_{xy}\sin2\theta$$

$$\tau_{x'y'} = -\frac{\sigma_x - \sigma_y}{2}\sin2\theta - \tau_{xy}\cos2\theta$$

These equations permit the calculation of σ and τ values for any desired set of axes if the values are known for any given set of axes. In particular, if σ is chosen properly, $\tau_{x'y'}$ can be made to vanish and normal stresses only will remain. The set of axes for which this is true are called the *principal axes* of stress and the corresponding σ's are called the *principal stresses*. All points within static beds of powders can be characterized by principal axes and stresses that will, in general, vary from point to point throughout the bed. The principal axes do not correspond necessarily to the orientation of the walls of the powder container.

These concepts can be extended to three dimensions. Thus, it is possible to find a set of three mutually perpendicular planes, on which there are no shear stresses acting, for each location within the powder. The normals to these planes are the principal axes. It also is possible to find a set of planes for which the shear stresses are a maximum and the normal stresses are equal. The associated axes are called the axes of maximum shear. These two sets of axes are important because they represent directions of bed failure were it to occur.

The relationships between stresses, as functions of θ, can be illustrated and determined graphically. Figure 37-18 is an example of a Mohr's circle diagram for stress. Such diagrams are based on the stress equations. This can be seen by comparing Figure 37-18 with the equations, noting the relationships of the stresses of θ. A Mohr diagram can be constructed for any point within the powder, permitting stresses to be resolved graphically into normal and shear components for any arbitrary choice of axes.

Steps in constructing a diagram are

1. Plot the center of the circle, p, on the σ axis at the average normal stress, $(\sigma_x + \sigma_y)/$
2. Plot point x and y with coordinates (σ_x, τ_{xy}) and (σ_y, τ_{xy}), respectively. Note that these three points lie on a diameter of the circle.
3. Draw a circle with its center at p and passing through points x and y.
4. Locate the $x'y'$ diameter using the angle 2θ.
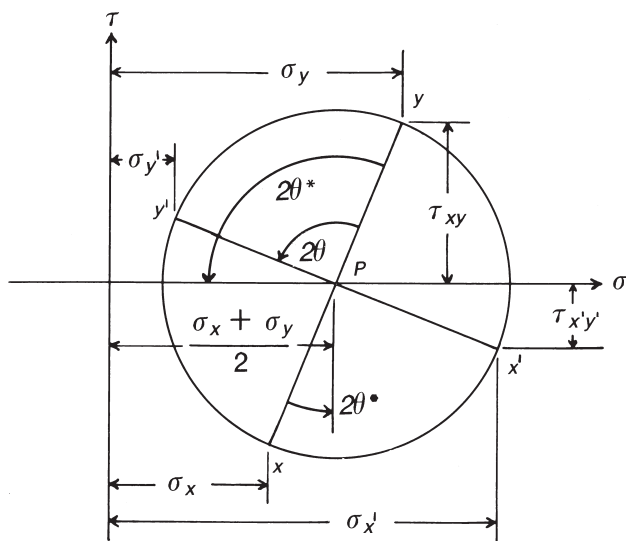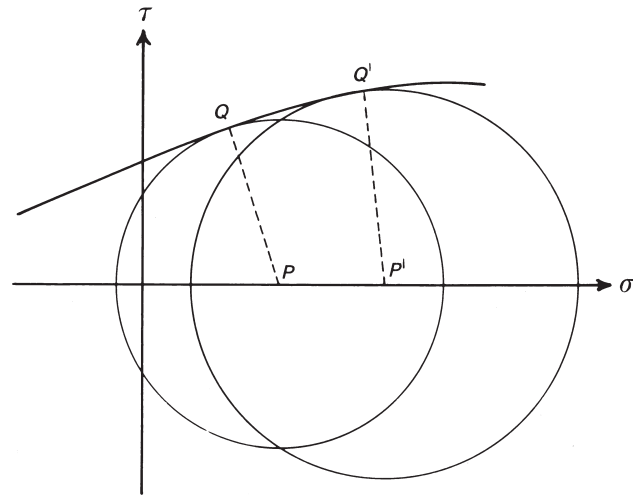
Figure 37-18.

Figure 37-19.

The stress components corresponding to the new axes can be read off the graph. Both $\sigma_{x'}$ and $\sigma_{y'}$ are read off the same axes on the graph because both are normal stresses.

For the particular case in Figure 37-19, the principal axes lie at an angle of θ* to the original axes. The axes of maximum shear stress lie at an angle of θ⁻ from the original axes because the xy line corresponding to maximum shear is perpendicular to the σ axis. Depending on the state of the powder, it is possible to have negative σ values, where the Mohr circle passes to the left of the τ axis.

The application of stress normal to a plane of shear influences the shear stress at which the powder fails. Because of this, a given powder will fail at various combinations of normal and shear stresses. These combinations can be expressed graphically by a line in the σ, τ plane that separates regions on the graph at which the powder either flows or is stable.

This is shown in Figure 37-19 for a typical powder. Various powders will display curves that uniquely define their failure characteristics. Each point on such a curve corresponds to a σ, τ combination at which failure occurs and can be analyzed by constructing a Mohr circle that passes through the point and is centered on the intersection of a line perpendicular to the point q and the σ axis. An example is shown in Figure 37-19.

BULK PROPERTIES

In addition to the angles of repose and friction that reflect bulk behavior, tensile and shear strength and dilatancy are of interest. Tensile strength is measured by forming a powder bed on a roughened and split plate. Half of the plate is laterally movable and the force necessary to rupture the bed by pulling the plate halves apart, minus sliding plate friction corrections, represents the bed tensile strength. Various methods of applying force to the movable plate are used, including tipping the plate from the horizontal and allowing it to react to gravity by rolling on steel balls.

Shear strength is determined from the force necessary to shear horizontally a bed of known cross-section. The Jenike shear cell is typical of those in use. It permits various loads to be applied normal to the plane of shear, whereby a shear failure locus can be determined. With the desired normal load applied, a steadily increasing shearing force is applied until failure occurs. These measurements are the basis for constructing powder-failure curves.

When packed powder beds are deformed, local expansion occurs along the failure planes, barring fracture of the particles themselves. This phenomenon is termed dilatancy and is a

direct consequence of the micromechanics of interparticulate movement. For one particle to move past another, it is necessary for it to move to the side in order to move forward when the particles are in an *interlocked* arrangement. Such arrangements predominate in packed beds with the consequence that the collective sideways movements in the failure zone produce bed expansion. Room for expansion therefore must be provided when packed beds are forced to flow.

Mixing of Powders

DEGREE OF HOMOGENEITY

Many mathematical expressions have been proposed and used to express the degree of homogeneity of powders composed of two or more components. For the most part, measures of mixture uniformity have been statistical and based on either the standard deviation or variance of the composition from its mean value. It should be recognized that these indices of mixing are scalar quantities and are incapable of uniquely describing the composition profile of a given powder bed. A practical definition of mixing uniformity should be selected to relate as closely as possible to the desired properties of the mix. The manner in which samples are taken (number, size, location of samples, and method of sampling) largely determines the validity and interpretation of the derived index[7].

The standard deviation is presented here as a representative index. It can be estimated solely from a set of n samples. If sample number i has composition x_i, and all samples are of uniform size, the sample standard deviation is defined in the usual way as

$$s = \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2/(n-1)}$$

where $\overline{x}$ is the mean composition estimated from the samples alone.

In sampling a bed, there should be assurance that the bed is sampled uniformly over its entirety. This can be done either by use of a *sampling thief* designed to probe the bed and collect samples at selected points or serially as the powder is discharged from the mixer.

The *scale of scrutiny* at which the powder is examined for uniformity is determined by the sample size. This should be chosen based on the ultimate use of the powder. For a tablet or capsule formulation, the appropriate sample size is that of the dosage form.

Two important concepts related to mixing uniformity have been described by Danckwerts as the scale and intensity of segregation. Assuming that zones having uniform but differing compositions exist in a powder bed, the scale of segregation is a function of the size of the zones. The intensity of segregation is, in turn, a function of the composition differences among zones. Generally, the process of mixing tends to reduce the intensity of segregation, whereas the scale of segregation passes through a minimum.

MECHANISMS OF MIXING AND SEGREGATION

Three primary mechanisms are responsible for mixing:

- Convective movement of relatively large portions of the bed.
- Shear failure, which primarily reduces the scale of segregation.
- Diffusive movement of individual particles.

Most efficient mixers operate to induce mixing by all three mechanisms. Thus, mixing can be considered to be a random shuffling-type operation involving both large and small particle groups and even individual particles. However, it should be noted that the use of random motion to achieve random distribution assumes that no other factors influence this distribution. This is rarely if ever the case in practice. Instead, a variety of properties of the powders being mixed influence this approach to complete randomness. Stickiness or slipperiness of particles must be considered, among other factors. As might be expected, the stickier the material, the less readily it mixes and demixes. Electrostatic forces on the particle surface also can produce marked effects on the mixing process, and in fact may produce sufficient particle-particle repulsion to make random mixing impossible.

By enabling particles to undergo movement relative to each other, mixers also provide the conditions necessary for segregation to occur. Any manipulation of a powder bed for purposes of conveying, discharging from a hopper, and so on provides the opportunity for segregation. Thus, many of the so-called mechanisms of segregation are actually conditions under which segregation can happen.

The segregation that occurs in free-flowing solids usually does so as a result of differences in particle size and, to a lesser extent, to differences in particle density and shape. The circumstances leading to segregation can be generalized from a fundamental physical standpoint. The necessary and sufficient conditions for segregation to occur are

1. Various mixture components exhibit mobilities for interparticulate movement that differ.
2. The mixture experiences either a field that exerts a directional motive force on the particles, or a gradient in a mechanism capable of inducing or modifying interparticulate movement.

The combination of these conditions results in asymmetric particle migrations and leads to segregation.

RATES OF MIXING AND SEGREGATION

Rate expressions analogous to those of chemical kinetics can be derived using any of the various indices of mixing as time-dependent variables. When this is done, it usually is found that mixing follows a first-order approach to an equilibrium state of mixedness. More recently, mixing has been described as a stochastic process (by means of stationary and nonstationary Markov chains) in which the probabilities of particle movement from place to place in the bed are determined. When applied to a mixer, this approach is capable of indicating zones of greater and lesser mixing intensity.

LARGE-SCALE MIXING EQUIPMENT

The ideal mixer should produce a complete blend rapidly with as gentle as possible a mixing action to avoid product damage. It should be cleaned and discharged easily, be dust-tight, and require low maintenance and low power consumption. All of these assets generally are not found in any single piece of equipment, thus requiring some compromise in the selection of a mixer.

Rotating-Shell Mixers—The drum-type, cubical-shaped, double-cone, and twin-shell blenders are all examples of this class of mixers. Drum-type blenders, with their axis of rotation horizontal to the center of the drum, are used quite commonly. These, however, suffer from poor crossflow along the axis. The addition of baffles or inclining the drum on its axis increases crossflow and improves the mixing action.

Cubical-and polyhedron-shaped blenders with the rotating axis set at various angles also are available. However, in the latter, because of their flat surfaces, the powder is subjected more to a sliding than a rolling action, a motion that is not conducive to the most efficient mixing.

Double-cone blenders, an important class of rotating-shell or tumbling mixers, were developed in an attempt to overcome some of the shortcomings of the previously discussed mixers. Here, the mixing pattern provides a good crossflow with a rolling rather than a sliding motion. Normally, no baffles are required, so cleaning is simplified. The twin-shell blender is another important tumbling-type blender. It combines the
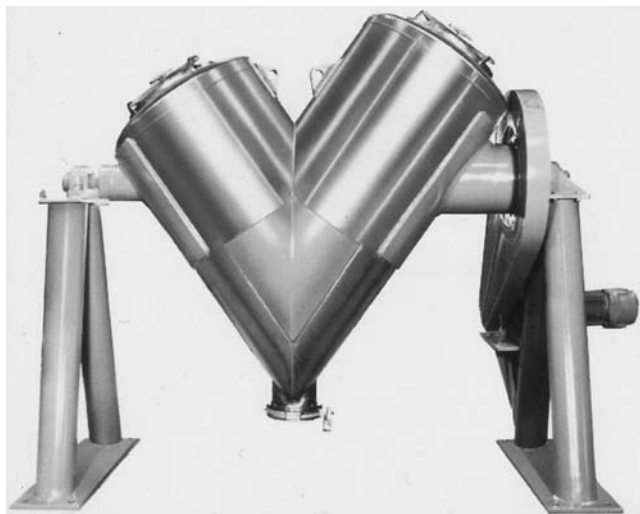
Figure 37-20. Cross-flow twin-shell blender (courtesy, Patterson-Kelley).

efficiency of the inclined drum-type with the intermixing that occurs when two such mixers combine their flow.

The Cross-Flow blender (*Patterson-Kelley*) (Fig 37-20) is an example of a twin-shell blender. The uneven length of each shell in this blender provides additional mixing action when the powder bed recombines during each revolution of the blender. The Zig-Zag blender, an extension of the twin-shell blender, provides efficient continuous precision blending.

Fixed-Shell Mixers—The ribbon mixer, one of the oldest mechanical solid–solid blending devices, exemplifies this type of mixer. It consists of a relatively long troughlike shell with a semicircular bottom. The shell is fitted with a shaft on which are mounted spiral ribbons, paddles, or helical screws, alone or in combination. These mixing blades produce a continuous cutting and shuffling of the charge by circulating the powder from end to end of the trough as well as rotationally. The shearing action that develops between the moving blade and the trough serves to break down powder agglomerates. However, ribbon mixers are not precision blenders; in addition, they suffer from the disadvantage of being more difficult to clean than the tumbler-type blenders and of having a higher power requirement.

Sigma-Blade and Planetary Paddle Mixers—Sigma-blade and planetary paddle mixers also are used for solid–solid blending, although most generally as a step prior to the introduction of liquids. Mixers with high-speed impeller blades set into the bottom of a vertical or cylindrical shell have been shown to be very efficient blenders. This type, in addition to its ability to produce precise blends, serves also to break down agglomerates rapidly. The mechanical heat buildup produced within the powder mix and the relatively high power requirement are often drawbacks to the use of this type of mixer; however; the shorter time interval necessary to achieve a satisfactory blend may offset these factors.

Vertical Impeller Mixers—Vertical impeller mixers, which have the advantage of requiring little floor space, employ a screw-type impeller that constantly overturns the batch (Fig 37-21). The fluidized mixer is a modification of the vertical impeller type. The impeller is replaced by a rapidly moving stream of air fed into the bottom of the shell. The body of the powder is fluidized, and mixing is accomplished by circulation and overtumbling in the bed (Fig 37-22). Generally, when precision solid–solid blending is required, the rotating twin-shell or the double-cone–type blenders are recommended.

Motionless Mixers—These are in-line continuous processing devices with no moving parts. They consist of a series of fixed flow-twisting or flow-splitting elements. The Blendex (*Ross & Son*), designed for blending of free-flowing solids, is constructed to operate in a vertical plane. Four pipes inter-

connect with successive tetrahedral chambers, the number of chambers needed depending on the quality of mix desired. The powders enter the mixer from overhead hoppers and free-fall through the mixer and are mixed by what is described as Interfacial Surface Generation. For two input streams entering this mixer the number of layers, L, emerging from each of the successive chambers, C, is $L = 2(4)^C$. Thus, for 10 chambers over 2 million layers are generated. This type provides efficient batch or continuous mixing for a wide variety of solids without particle-size reduction or heat generation with essentially no maintenance. Units are available to mix quantities ranging from 100 to 5000 lb/hour.

SMALL-SCALE MIXING EQUIPMENT

The pharmacist most generally employs the mortar and pestle for the small-scale mixing usually required for prescription compounding. However, spatulas and sieves also may be used on occasion. The mortar and pestle method combines comminution and mixing in a single operation. Thus, it is particularly useful where some degree of particle-size reduction as well as mixing is required, as in the case of mixtures of crystalline material.

The blending of powders with a spatula on a tile or paper, or spatulation, is used sometimes for small quantities of powders, often as an auxiliary blending technique or when the compaction produced by the mortar and pestle technique is
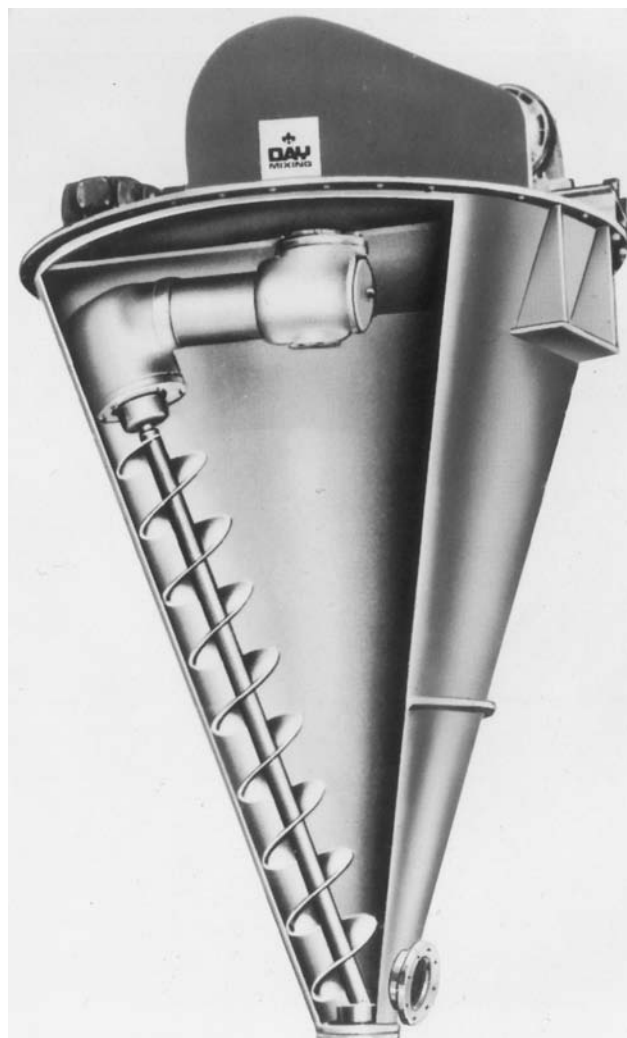

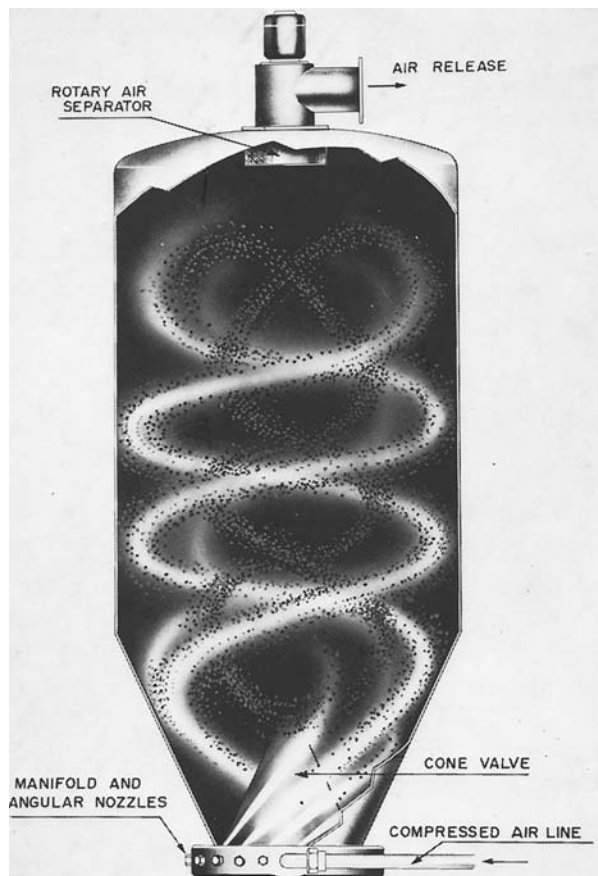**Figure 37-21.** Cutaway view of the Mark II Mixer (courtesy, JH Day).

Figure 37-22. Fluidized air mixer (courtesy, Sprout, Waldron).

undesirable. Spatulation is a relatively inefficient method of mixing, thus principles of geometric dilution must be employed. Spatulation is rarely used to prepare a finished dosage form.

Sieving usually is employed as a pre- or post-mixing method to reduce loosely held agglomerates and to increase the overall effectiveness of a blending process. When used alone as a solid–solid blending technique, several passes through the sieve are required to produce a reasonably homogeneous mix.

Storage and Flow

FLOW PATTERNS

Discharge of powders from large-scale mixers, storage, bins or machine-fed hoppers primarily generates flow in the form of shear failure—the powder behaves in a manner analogous to a viscous liquid in laminar flow. The analogy ends at that point, because conditions are then present in the powder bed conducive to segregation. The overall pattern of discharge from a bin takes the form of either funnel flow or mass flow. Bin-design characteristics, which take into account the powder's angles of slide and internal friction and its yield locus in terms of normal and shear stresses, determine which flow pattern will occur.

In funnel flow the powder moves in a column down the center of the bin toward the exit orifice at the bottom. Material surrounding this relatively rapidly moving core remains stationary or is drawn slowly into the core, which is fed primarily from the top where powder moves to the center and then down in the manner of a funnel.

The powder in a mass-flow bin moves downward toward the orifice as a coherent mass. When it reaches the tapered section of the bin leading to the orifice, it is compressed and flows in shear analogous to a plastic mass being compressed. This type

of bin is advantageous for use with powders having a strong tendency to segregate.

The rate of discharge from a hopper varies as a function of the cube of the orifice diameter and is nearly independent of the height of the bed. An arch forms over the orifice that in effect is a boundary between material in essentially free-fall and material in the closely packed condition of the powder bed. The rate of mass transport across this constantly renewed surface determines the rate of orifice flow. It has been shown that flow can be increased substantially if gas is pumped through the bed and across the orifice in the direction of the solids flow. Flow conditioners, an important means of improving flow, are discussed in Chapter 20.

PNEUMATIC TRANSPORT

The pneumatic transport of powders is of interest because it can be used to mix powders at the same time as they are being conveyed. The method consists of propelling a solids–gas mixture along a conduit via a gas pressure drop. The solids are held in suspension by the turbulence of the gas stream. At low-solids concentrations, where the particles are relatively small, the solids are dispersed uniformly over the pipe cross-section. However, at higher solids content or with larger particles, some stratification will occur in a horizontal pipe and solids will settle out if the pipe is overloaded.

Gas flow must be turbulent so as to suspend the solids; however, the solids behave as in laminar flow. Slippage between gas and solid occurs, particularly in vertical pipes; consequently, gas and solids flow rates are not in proportion to flow-stream composition. Further, smaller and less dense particles flow more rapidly than large and dense material and a chromatographic-like separation occurs. This is not a problem, however, once steady state is achieved. Because of the industrial importance of this process in many fields it has been investigated extensively and a number of useful theoretical and empirical expressions have been derived and may be used to predict conditions necessary for satisfactory pneumatic transport.

POWDERS AS A DOSAGE FORM

Historically, powders represent one of the oldest dosage forms. They are a natural outgrowth of the attempt to prepare crude drugs and other natural products in a more conveniently administered form. However, with declining use of crude drugs and increasing use of many highly potent compounds, powders as a dosage form have been replaced largely by capsules and tablets.

In certain situations, powders possess advantages and thus still represent a portion (although small) of the solid dosage forms currently being employed. These advantages are flexibility in compounding and relatively good chemical stability. The chief disadvantages of powders as a dosage form are they are time-consuming to prepare and they are not well suited for dispensing the many unpleasant-tasting, hygroscopic, or deliquescent drugs.

Bulk powders have another serious disadvantage when compared with divided and individually weighed powders: inaccuracy of dose. The dose is influenced by many factors, including size of measuring spoon, density of powder, humidity, degree of settling, fluffiness due to agitation, and personal judgment. Not only do patients measure varying amounts of powder when using the same spoon, but they often select one differing in size from that specified by their physician.

EXTEMPORANEOUS TECHNIQUES

In both the manufacturing and extemporaneous preparation of powders, the general techniques of weighing, measuring,

sifting, and mixing, as described previously, are applied. However, the following procedures should receive special attention.

- Use of geometric dilution for the incorporation of small amounts of potent drugs.
- Reduction of particle size of all ingredients to the same range to prevent stratification of large and small particles.
- Sieving when necessary to achieve mixing or reduction of agglomerates, especially in the preparation of dusting powders or powders into which liquids have been incorporated.
- Heavy trituration, when applicable, to reduce the bulkiness of a powder.
- Protection against humidity, air oxidation, and loss of volatile ingredients.

Powders are prepared most commonly either as divided powders and bulk powders, which are mixed with water or other suitable material prior to administration, or as dusting powders, which are applied locally. They also may be prepared as dentifrices, products for reconstitution, insufflations, aerosols, and other miscellaneous products.

The manually operated procedures usually employed by the pharmacist today are *trituration, pulverization by intervention,* and *levigation.*

Trituration—This term refers to the process of reducing substances to fine particles by rubbing them in a mortar with a pestle. The term also designates the process whereby a mixture of fine powders is intimately mixed in a mortar. The circular mixing motion of the pestle on the powders contained in a mortar blends the powders and also breaks up their soft aggregates. By means of the application of pressure on the pestle, crushing or grinding also can be effected. When granular or crystalline materials are to be incorporated into a powdered product, these materials are comminuted individually and then blended together in the mortar.

Pulverization by Intervention—This is the process of reducing the state of subdivision of solids with the aid of an additional material that can be removed easily after the pulverization has been completed. This technique often is applied to substances that are gummy and tend to reagglomerate or that resist grinding. A prime example is camphor, which cannot be pulverized easily by trituration because of its gummy properties; however, on the addition of a small amount of alcohol or other volatile solvent, this compound can be reduced readily to a fine powder. Similarly, iodine crystals may be comminuted with the aid of a small quantity of ether. In both instances the solvent is permitted to evaporate and the powdered material is recovered.

Levigation—In this process a paste is first formed by the addition of a suitable nonsolvent to the solid material. Particle-size reduction then is accomplished by rubbing the paste in a mortar with a pestle or on an ointment slab using a spatula. Levigation generally is used by the pharmacist to incorporate solids into dermatological and ophthalmic ointments and suspensions.

THE MORTAR AND PESTLE

The mortar and pestle are the most frequently used utensils in small-scale comminution. Mortars made of various materials and in diverse shapes are available; although these often are used interchangeably, the different kinds of mortars have specific utility in preparing or grinding different materials.

Modern mortars and pestles are prepared usually from Wedgwood ware, porcelain, or glass. Although pharmacists often use different mortars interchangeably, each type has a preferential range of utility.

Glass mortars are designed primarily for use in preparing solutions and suspensions of chemical materials in a liquid. They also are suitable for preparing ointments which require the reduction of soft aggregates of powdered materials or the incorporation of relatively large amounts of liquid. Glass also has the advantage of being comparatively nonporous and of not staining easily and thus is particularly useful when substances such as flavoring oils or highly colored substances are used. Glass cannot be used for comminuting hard solids.

Wedgwood mortars are suited well for comminution of crystalline solids or for the reduction in particle size of most materials used in modern prescription practice. They are capable of adequately powdering most substances that are available only as crystals or hard lumps. However, Wedgwood is relatively porous and will stain quite easily. A Wedgwood mortar is available with a roughened interior, which aids in the comminution process but which requires meticulous care in washing because particles of the drugs may be trapped in the rough surface and cause contamination of materials subsequently comminuted in the mortar.

Porcelain mortars are very similar to Wedgwood, except that the exterior surface of the former is usually glazed and thus less porous. Porcelain mortars may be used for comminution of soft aggregates or crystals but more generally are used for blending powders of approximately uniform particle size.

Pestles are made of the same material as the mortar. Pestles for Wedgwood or porcelain mortars are available with hard rubber or wooden handles screwed into the head of the pestle. Also available are one-piece Wedgwood pestles. Pestles made entirely of porcelain are objectionable, because they are broken easily.

Pestles and mortars should not be interchanged. The efficiency of the grinding or mixing operation depends largely on a maximum contact between the surfaces of the head of the pestle and the interior of the mortar. The pestle should have as much bearing on the interior surface of the mortar as its size will permit. A pestle that does not *fit* the mortar will result in a waste of labor.

Divided Powders

Divided powders (*chartula* or *chartulae*) are dispensed in the form of individual doses and generally are dispensed in papers, properly folded. They also may be dispensed in metal foil, small heat-sealed plastic bags, or other containers.

DIVIDING POWDERS

After weighing, comminuting, and mixing the ingredients, the powders must be divided accurately into the prescribed number of doses. To achieve accuracy consistent with the other steps in the preparation, *each dose should be weighed individually* and transferred to a powder paper. Following completion of this step, the powder papers are folded.

FOLDING POWDERS

The operations of folding powder papers are illustrated in Figure 37-23. Care in making the several folds, and experience gained by repetition, are necessary to obtain uniformity when the powders finally are placed in the box for dispensing. Deviation from any of the three main folds will result in powders of varying height being formed, and variations in the folded ends likewise will be noticeable when the powders are placed side by side.

PACKAGING DIVIDED POWDERS

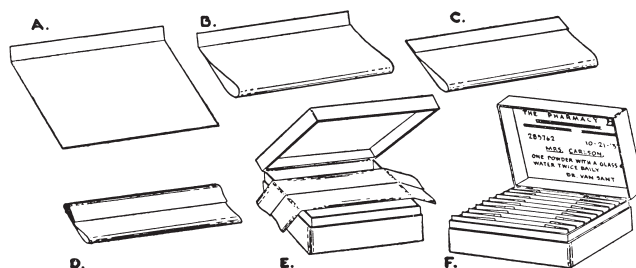Specially manufactured paper and boxes are available for dispensing divided powders.

Figure 37-23. Folding powder papers.

Powder Papers—Four basic types of powder papers are available.

1. Vegetable parchment, a thin, semiopaque, moisture-resistant paper.
2. White bond, an opaque paper with no moisture-resistant properties.
3. Glassine, a glazed, transparent, moisture-resistant paper.
4. Waxed, a transparent waterproof paper.

Hygroscopic and volatile drugs can be protected best by using a waxed paper, double-wrapped with a bond paper to improve the appearance of the completed powder. Parchment and glassine papers offer limited protection for these drugs.

A variety of sizes of powder papers are available. The selection of the proper size depends on the bulk of each dose and the dimensions of the powder box required to hold the number of doses prescribed.

Powder Boxes—Various types of boxes are supplied in several sizes for dispensing divided powders. The hinged-shoulder box shown in Figure 37-23F is the most popular; these have the advantage of preventing the switching of lids with the directions for use when several boxes of the same size are in the same home. The prescription label may be pasted directly on top of the lid or inside the lid. In the latter case, the name of the pharmacy is lithographed on top of the lid.

SPECIAL PROBLEMS

The incorporation of volatile substances, eutectic mixtures, liquids, and hygroscopic or deliquescent substances into powders presents problems that require special treatment.

VOLATILE SUBSTANCES

The loss of camphor, menthol, and essential oils by volatilization when incorporated into powders may be prevented or retarded by use of heat-sealed plastic bags or by double wrapping with a waxed or glassine paper inside of a bond paper.

EUTECTIC MIXTURES

Liquids result from the combination of phenol, camphor, menthol, thymol, antipyrine, phenacetin, acetanilid, aspirin, salol, and related compounds at ordinary temperatures. These so-called eutectic mixtures may be incorporated into powders by addition of an inert diluent. Magnesium carbonate or light magnesium oxide are commonly used, effective diluents for this purpose, although kaolin, starch, bentonite, and other absorbents have been recommended. Silicic acid prevents eutexia with aspirin, phenyl salicylate, and other troublesome compounds; incorporation of about 20% silicic acid (particle size, 50 μm) prevented liquefaction even under the compression pressures required to form tablets.

In handling this problem, each eutectic compound should be mixed first with a portion of the diluent and gently blended together, preferably with a spatula on a sheet of paper. Generally, an amount of diluent equal to the eutectic compounds is sufficient to prevent liquefaction for about 2 weeks. Deliberate forcing of the formation of the liquid state, by direct trituration, followed by absorption of the moist mass, also will overcome this problem. This technique requires use of more diluent than previously mentioned methods but offers the advantage of extended product stability. Thus, the technique is useful for dispensing a large number of doses that normally would not be consumed over a period of 1 or 2 weeks.

LIQUIDS

In small amounts, liquids may be incorporated into divided powders. Magnesium carbonate, starch, or lactose may be added to increase the absorbability of the powders if necessary. When the liquid is a solvent for a nonvolatile heat-stable compound, it may be evaporated gently on a water bath. Lactose may be added during the course of the evaporation to increase the rate of solvent loss by increasing the surface area. Some fluidextracts and tinctures may be treated in this manner, although the use of an equivalent amount of a powdered extract, when available, is a more desirable technique.

HYGROSCOPIC AND DELIQUESCENT SUBSTANCES

Substances that become moist because of affinity for moisture in the air may be prepared as divided powders by adding inert diluents. Double-wrapping is desirable for further protection. Extremely deliquescent compounds cannot be prepared satisfactorily as powders.

BULK POWDERS

Bulk powders may be classified as oral powders, dentifrices, douche powders, dusting powders, insufflations, and triturations.

ORAL POWDERS

Oral powders generally are supplied as *finely divided powders* or *effervescent granules.* The finely divided powders are intended to be suspended or dissolved in water or mixed with soft foods such as applesauce prior to administration. Antacids and laxative powders frequently are administered in this form.

Effervescent granules contain sodium bicarbonate and either citric acid, tartaric acid, or sodium biphosphate in addition to the active ingredients. On solution in water, carbon dioxide is released as a result of the acid–base reaction. The effervescence from the release of the carbon dioxide serves to mask the taste of salty or bitter medications.

Granulation generally is accomplished by producing a moist mass, forcing it through a coarse sieve and drying it in an oven. The moisture necessary for massing the materials is obtained readily by heating them sufficiently to drive off the water of hydration from the uneffloresced citric acid. The completed product must be dispensed in tightly closed glass containers to protect it against the humidity of the air.

Effervescent powders may be prepared also by adding small amounts of water to the dry salts to obtain a workable mass. The mass is dried and ground to yield the powder or granule. Care must be used in this procedure to ensure that the reaction that occurs in the presence of water does not proceed too far before it is stopped by the drying process. Should this happen, the effervescent properties of the product will be destroyed.

Other preparative techniques have been reported for effervescent powders such as a fluidized-bed procedure in which the powders are blended and then suspended in a stream of air in a Wurster chamber. Water is sprayed into the chamber, resulting in a slight reaction and an expansion of the particles to form granules ranging in size from 10- to 30-mesh. This approach apparently offers a number of advantages over the older techniques. The extent of reaction and particle size are controlled during the manufacture. A drying oven, trays, or even grinding devices are not required. Furthermore, the technique lends itself to a continuous as well as a batch operation.

The heat generated from the blending and mixing operation also has been used to mass the powders by causing the release of the water of hydration from the citric acid. The massed materials can be dried and sieved through a coarse sieve. This technique thus eliminates the need of an external heat source or a granulating solution.

DENTIFRICES

Dentifrices may be prepared in the form of a bulk powder, generally containing a soap or detergent, mild abrasive, and an anticariogenic agent.

DOUCHE POWDERS

Douche powders are completely soluble and are intended to be dissolved in water prior to use as antiseptics or cleansing agents for a body cavity. They most commonly are intended for vaginal use, although they may be formulated for nasal, otic, or ophthalmic use. Generally, because aromatic oils are included in these powders, they are passed through a No 40 or 60 sieve to eliminate agglomeration and ensure complete mixing. Dispensing in wide-mouth glass jars serves to protect against loss of volatile materials and permits easy access by the patient. Bulk-powder boxes may be used for dispensing douche powders, although glass containers are preferred because of the protection afforded by these containers against air and moisture.

DUSTING POWDERS

Dusting powders are locally applied nontoxic preparations that are intended to have no systemic action. They are applied to various parts of the body as lubricants, protectives, absorbents, antiseptics, antipruritics, antibromhidrosis agents, astringents, and antiperspirants. Dusting powders always should be dispensed in a very fine state of subdivision to enhance effectiveness and minimize irritation. When necessary, they may be micronized or passed through a No 80 or 100 sieve.

Extemporaneously prepared dusting powders should be dispensed in sifter-top packages. Commercial dusting powders are available in sifter-top containers or pressure aerosols. The latter, while generally more expensive than the other containers, offer the advantage of protection from air, moisture, and contamination, as well as convenience of application. Foot powders and talcum powders are currently available as pressure aerosols.

Although in most cases dusting powders are considered nontoxic, the absorption of boric acid through large areas of abraded skin has caused toxic reactions in infants. Accidental inhalation of zinc stearate powder has led to pulmonary inflammation of the lungs of infants. The pharmacist should be aware of the possible dangers when the patient uses these compounds as well as other externally applied products. See also Chapter 65.

INSUFFLATIONS

Insufflations are finely divided powders introduced into body cavities such as the ears, nose, throat, tooth sockets, and vagina. An insufflator (powder blower) usually is employed to administer these products. However, the difficulty in obtaining a uniform dose has restricted their general use.

Specialized equipment has been developed for the administration of micronized powders of relatively potent drugs. The Norisodrine Sulfate Aerohaler Cartridge (*Abbott*) is an example. In the use of this Aerohaler, inhalation by the patient causes a small ball to strike a cartridge containing the drug. The force of the ball shakes the proper amount of the powder free, permitting its inhalation. Another device, the Spinhaler turbo-inhaler (*Fisons*), is a propeller-driven device designed to deposit a mixture of lactose and micronized cromolyn sodium into the lung as an aid in the management of bronchial asthma.

Pressure aerosols also have been employed as a means of administering insufflations, especially for potent drugs. This method offers the advantage of excellent control of dose, through metered valves, as well as product protection.

TRITURATIONS

Triturations are dilutions of potent powdered drugs, prepared by intimately mixing them with a suitable diluent in a definite proportion by weight. They were at one time official as 1 to 10 dilutions. The pharmacist sometimes prepares triturations of poisonous substances such as atropine in a convenient concentration using lactose as the diluent, for use at the prescription counter. These medicinal substances are weighed more accurately and conveniently by using this method.

The correct procedure for preparing such triturations or any similar dilution of a potent powder medicament, to ensure uniform distribution of the latter, is

1. Reduce the drug to a moderately fine powder in a mortar.
2. Add about an equal amount of diluent and mix well by thorough trituration in the mortar.
3. Successively add portions of diluent, triturating after each addition, until the entire quantity of diluent has been incorporated.

Under no circumstance should the entire quantity of diluent be added at once to the drug that is to be diluted in the expectation that uniform dispersion of the latter will be more expeditiously achieved on brief trituration of the mixture.

REFERENCES

1. Parrott EL. In Lachman L, et al. *The Theory and Practice of Industrial Pharmacy,* 3rd ed. Philadelphia: Lea & Febiger, 1986, p 32.
2. Perry RH, et al. *Chemical Engineers' Handbook,* 7th ed. New York: McGraw-Hill, 1997: 8–8.
3. Byers JE, Peck GE. *Drug Dev Ind Pharm* 1990; 16(11): 1761–1779.
4. Pilcher JM, et al. *Proc Chem Spec Mfrs Assoc Ann Mtg* 1956; 66.
5. Tillotson D. *Aerosol Age* 1958; 3(5): 41.
6. Rawle A. *Adv. Colour Sci Tech* 2002; 5(1):1–12.
7. Muzzio FJ, et al. *Int J Pharm* 2003; 250:51–64.

BIBLIOGRAPHY

Alderborn G, Nystrom C., eds. *Pharmaceutical Powder Compaction Technology*. New York: Marcel Dekker, 1996.
Allen T. *Particle Size Measurement,* 5th ed. London: Chapman & Hall, 1997.
Brittain HG, ed. *Physical Characterization of Pharmaceutical Solids*, New York: Marcel Dekker, 1995.
Carstensen JT. *Advanced Pharmaceutical Solids*, New York: Marcel Dekker, 2000.
Hickey AJ and Ganderton D. *Pharmaceutical Process Engineering*, New York: Marcel Dekker, 2001.
Levin M, ed. Pharmaceutical Process Scale-Up, New York: Marcel Dekker, 2002.
Martin AN, et al. *Physical Pharmacy,* 4th ed. Philadelphia: Lea & Febiger, 1993.
Venables HJ, Wells JI. Powder Mixing, *Drug Dev Ind Pharm* 2001; 27(7): 599–612.

Property-Based Drug Design and Preformulation

Howard Y Ando, PhD
Galen W Radebaugh, PhD

The discovery and development of new chemical entities (NCEs) into stable, bioavailable, marketable drug products is a long, but rewarding process. Due to the tremendous cost of developing a NCE, and industry's need to enhance productivity, it is desirable to create NCEs that have suitable physical-chemical properties, rather than compensate for deficiencies solely by the formulation process. Hence, property-based design can enhance the likelihood a NCE will have the desired physical-chemical that will facilitate its ability to be developed into a stable, bioavailable dosage form. Even so, well-designed preformulation studies are necessary to fully characterize molecules during the discovery and development process so that NCEs have the appropriate properties, and there is an understanding of the deficiencies that must be overcome by the formulation process. This chapter provides guidance that will facilitate property-based design and the supporting preformulation studies necessary to direct formulation efforts to give NCEs the highest possibility of success.

EVOLUTION OF THE DRUG DISCOVERY PROCESS

The need for property-based design follows from the natural evolution of a research and development process that seeks to become more efficient. The growth and decline of markets and sectors is a natural process that applies to every life structure whether it is the universe, an individual, or a market sector. All have a sigmoidal curve with periods of vulnerability, growth, and decline. For the pharmaceutical new chemical entity (NCE) sector, this is shown in Figure 38-1 as NCE-1. Of course, the declining phase is of major concern and usually is seen only in retrospect. However, Charles Handy has pointed out that given enough foresight, organizations can renew themselves by changing their operational paradigm.[1] Ideally, they would initiate and build the basis for this change during the α phase (shown in Fig 38-1). If successful, they could then initiate the hypothetical second curve, labeled NCE-2 in Fig 38-1. What then are the causes for the aging of the NCE-1 cycle, and what will fuel the initiation and growth of the hypothetical NCE-2 cycle paradigm? The relevance of property-based design in this context is discussed below.

GROWTH CYCLE DETERMINANTS

NCE Paradigms

The first growth epoch for the pharmaceutical development was driven by the application of physical-chemical principles to the design of dosage forms and delivery of NCEs. Physical chemistry provided scientists with a macroscopic, theoretical model, and as a young discipline, empirical experimentation predominated in the industrial design of dosage forms. Moreover, discovery and development phases occurred as separate and sequential phases. This was efficient and sufficient at the time, mainly because the targets were simpler. Evaluating the activity of new NCEs might involve bacterial cultures or perfused animal tissues. Testing for pharmacological activity in whole animals would then follow. Compounds that had poor development potential like limited aqueous solubility never showed any *in vivo* pharmacodynamic activity and were never advanced. In addition, indirect biomarkers were not needed because the physiological impact of an NCE could be readily measured and extrapolated from animals to humans (eg, blood pressure monitoring). However, new technological developments have caused the decline of this paradigm.

Advances in biotechnology fueled the second epoch starting in the 1980s because proteins could be synthesized from genetic information. Initially, bacteria and then mammalian cells were the source of these proteins. Such technology meant that these proteins could now be used as targets for discovery research. Individual receptors, enzymes, or transporters could now be synthesized in isolation from their parent tissue and could be used as surrogates for *in vivo* pharmacological activity. The banks of compounds that were accumulated during the first epoch, both in the academic and industrial setting, could now be screened for *in vitro* activity by high-speed robots.

The realization that a more integrated process of discovery was necessary became apparent only after a painful period. Early in this second epoch, a lot of energy was devoted to compounds that have been coined high affinity traps.[2] These are compounds that have very high *in vitro* activity but poor aqueous solubility. This occurred because of the needs of high throughput screening to automate the dispensing of compounds in a 96 well format. Because accurate and economical dispensing of powder is not possible, all reagents must be added as solutions. Liquid dispensing required a very general way to dissolve compounds. So the solution was to use small amounts of a very good, universal solvent, DMSO, that dissolved almost all organic compounds. The problem was that property-based factors like solubility and dissolution are not accounted for. Lipinski sounded the warning to the industry with his rule of five (RoF).[3] Subsequently, developmental scientists have put into place a number of high throughput physical property screens that could be used during the discovery phase; hence the realization of a need for property-based design. However, there are signs that this epoch may be reaching the end of its growth phase. DiMasi[4] has shown that the NCE-1 curve in Figure 38-1 for new INDs filings reached a plateau during the 1980s and has declined in the 1990s.
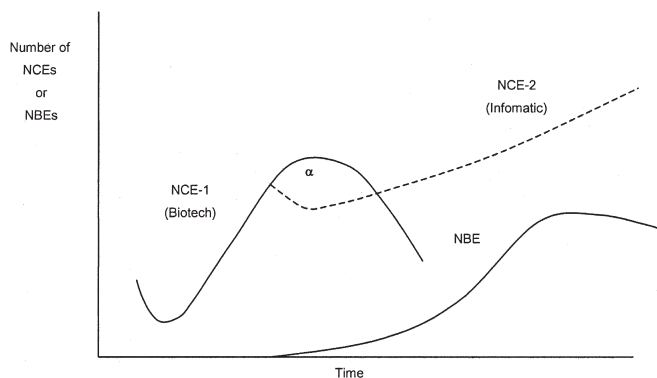
Figure 38-1. Charles Handy's sigmoidal growth curve. (From Handy C. The Age of Paradox. Cambridge, MA: Harvard Business School, 1995: 49-67. Copyright © 1995 by the Harvard Business School Publishing Corporation: all rights reserved.

Because the biotechnology paradigm may now be reaching the limits of its efficiency, it is proposed that a new paradigm (Informatics) will begin to evolve, taking advantage of an increased molecular understanding of the crystalline state and advances in the computational sciences, especially machine learning. The α phase of Figure 38-1 may be upon us. This new paradigm, NCE-2, will be driven by both technological opportunities, especially infomatics, and pharmaco-economic constraints.

Pharmaco-Economic Constraints

COST—In a recent white paper by IBM consultants, it was pointed out that the innovative driving force for drug development is rapidly shifting from the manufacturers and physicians to consumers, which in many cases are managed care organizations (MCO). One of the most important imperatives of this new consumer is the control of rising health care cost. With their control of formularies, MCOs will exert considerable influence in the future on the direction and limits of innovation.[5]

REGULATORY AND SAFETY—At the same time, regulatory agencies are requiring electronic filing requirements that in the short term considerably increase cost, but in the long term have the potential to speed review. In addition, because our understanding of side effects has increased substantially during the biotechnology epoch, self-imposed industry and regulatory requirements for NCEs have become much more stringent. For example, safety screens are now available for certain types of potentially fatal arrhythmias (torsades de pointes syndrome) that have been found to be associated with drug binding to potassium channels in the heart's conduction fibers. Chromosomal genotoxicity screens are also available that can detect a drug's interference with normal mitotic spindle and microtubule complex formations, or DNA strand breakage.[6] All of these new insights increase what is expected for a new NCE before it can be introduced into the marketplace. How then can costs be reduced as NCE regulatory requirements increase?

RISK MINIMIZATION—DiMasi has shown that the clinical approval rates from more recent IND filings has improved.[7] Apparently, better preclinical screening has increases the success rate. Since filtering out poor clinical candidates during the preclinical screening stage should be much cheaper than having clinical candidates fail, highly efficient screening should be justified. On the other hand, even if current preclinical screening is efficient in increasing the clinical success rate, apparently it does not add to productivity as measured by the decline in IND filings in the 1990s.[6] The substantial improvements that are needed to reduce both cost and risk and to initiate the Informatic NCE-2 curve in Figure 38-1 will most likely need the simultaneous improvements of a number of infomatic-based at point α.

Such improvements would include computational (a) activity-based design, (b) safety-based design, and (c) property-based design. If all of these elements could be highly accurate and applied at very early stages of discovery, fewer resources would be expended on nonproductive activities. In addition, if the number of potential opportunities both from the number of targets due to genomic opportunities and from increased property-based design possibilities can be achieved, then higher productivity should result.

Cost Reduction by Learning Before Doing

A model for the cost saving of such a paradigm has been carried out in the chemical development arena, but the concepts should hold for the property-design area as well. Today, when discovery chemists find a compound that has promising activity, additional amounts need to be made for further testing. Here the speed at which a chemical can be manufactured is critical. Usually, any route that will make the compound the quickest to synthesize on a small scale is chosen. If however, the compound continues to show potential, it has to be scaled up for even further testing. In his study, Pisano found that the two most important elements for reducing cost of manufacturing chemicals are: (a) the optimal synthetic route, and (b) telescoping successive unit operations. Of these two elements, finding the optimal synthetic route is the most important. If the company can effectively utilize its past experience to make the route determination earlier, then costs are reduced most effectively. Figure 38-2 shows the savings of this *learning before doing*.[8] One can imagine sometime in the not too distant future discovery chemists making decisions on which compounds to move forward based on all of the discovery criteria previously discussed but also on chemical synthesis scalability and optimum route design. Not only would the speed for making NCEs benefit, but also the long-term cost and efficiency of the entire chemical development organization.

In summary, the development cost can in theory be drastically reduced if computational design of property, activity, and safety can be accomplished. Such savings have the potential to alter the pharmaceutical industry's focus on blockbuster NCEs to potentially smaller but still lucrative markets. Accomplishment of this goal would most likely initiate the NCE-2 curve of Figure 38-1. The biotechnology arena is a good model. In Figure 38-1, the new biologic entities (NBEs) are seen to be growing as the NCEs are shown to be flat or peaked.[4]

INTEGRATION OF DISCOVERY AND DEVELOPMENT

As discussed, the pharmaceutical industry has evolved from a sequential organization where problems were passed on from discovery to development (epoch 1) to one in which both drug activity and physical properties are considered very early in discovery (epoch 2). The RoF was one of the early movements to
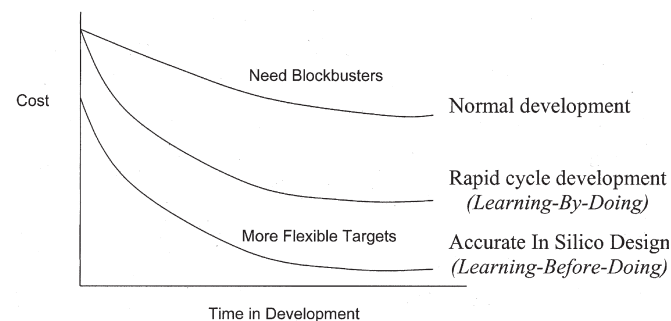
Figure 38-2. Cost savings by learning-before-doing.

foster integration of discovery and development. The ideal development of a NCE optimizes both "property-based" as well as "activity"-based design simultaneously. Continued improvements in efficiency will require that organizations be ready to adapt to new technologies and learnings. However, potential roadblocks to the integration of discovery and development efforts include high throughput (HT) decision-making, attrition, and the management of complexity.

HT Decision-Making

One of the attractive concepts for improving efficiency is that of successive screens. Currently, they come in two flavors, *in vitro* and computational to filter out poor drug candidates so resources are not wasted on unproductive activities.[9] The sequential paradigm

Discovery → Development

can now be replaced by the sequence

Discovery [design → synthesis] →
Selection [screen for activity → absorption → metabolism → toxicology] →
Development [formulation → animal pk testing → regulated toxicology → IND → initial clinical trials]

In essence, screens used in this manner are a way to simplify the complex process of discovering, selecting, and developing NCEs.

As efficient and useful as successive, hierarchical high throughput screens (HTSs) are for simplifying decision making, the question should be asked, "Have HTSs increased productivity?" As we alluded to under a previous section, productivity for IND filings (a measure of preclinical activity) has reached a plateau. This is most likely due to the use of successive filters in a decision-tree that then successively reduces future possibilities. If successive filters are employed, they could be prioritized so that earlier filters have higher quality. This would minimize the loss of potential opportunities.

Consider a situation of form selection in which scientists are trying to select the best molecule for development. In this multi-tiered approach, decision-making follows a progression of:

Hygroscopicity → thermal analysis & x-ray diffraction → accelerated solid-state stability

One impact of such decision-trees is that hygroscopic salts would rarely be developed (even if they have very advantageous bioavailability properties). If hygroscopicity were a property that prevented development, then any compound with this characteristic would be eliminated immediately. However, it is possible, with a good enough reason, to work with this situation.

Attrition

GAINS—Property-based screens have made tremendous gains over the last 5 years. This is due to the design of NCEs that have both activity and desirable physical properties such as solubility. These advances have been instrumental in reducing pharmacokinetic attrition during clinical trials.[7,10,11] On the other hand, more sophisticated technologies are needed to overcome low productivity problems associated with simple successive filters.

LOSSES—IMPACT OF FILTER IMPERFECTIONS—Reduced compound flow in the pipeline is a possible result of attrition filters. If these filters were perfect, this would not be a concern. Filters, however, hold back: (a) absolute negatives, (b) technical negatives, and (c) false negatives. Absolute negatives are compounds that are incompatible with the body. Consider, for example an insoluble, high affinity trap compound with a very high melting point (>240° C). Even if a pharmaceutical scientist were able to successfully formulate this compound for an intravenous formulation, it would most likely crystallize in the kidney. On the other hand, suppose water solubility was used as a filter. A technical negative that fails for adequate

water solubility, may still be biocompatible. A highly lipophilic compound with a melting point of 100° C would be a compound of this type. This compound may be deliverable by special formulations and has the potential to be a viable NCE from the property-design point of view. However, both of these compounds would be screened out if water solubility were used as an attrition filter. The final type of negative is a false negative in which the filter removes a perfectly viable compound.

To appreciate the impact of losing good compounds as false negatives and formulatable technical negatives, consider the following situation. Three filters A, B, and C are to be used in succession. To calculate their impact, assume that each has the following characteristics. Each will pass 50% of the positives correctly, will block correctly the 25% absolute negatives, but will also block 25% of compounds that are either false or technical negatives. For this battery of successive filters the throughput of positives is 12.5%. However, the correct throughput of positives and formulable compounds is 42%. Thus the pipeline possibilities were reduced unnecessarily by 236%. How many compounds are being filtered out that previously might have taken a considerable amount of time to develop but were developable? A key goal for property-design should be not to lose technical negatives that a company has the core competencies to develop rapidly.

PROPERTY-BASED DESIGN IN LEAD SELECTION

One of the keys for continuous improvement and moving into the Informatic α phase of Figure 38-1 is to make better use of existing data and to obtain higher quality data. In addition, the active participation of special groups that have domains of expertise is also needed. As we have seen, simple models can promote efficiency but more sophisticated refinements that take into account complexity are needed to increase productivity.

As an example, one area of extreme complexity is understanding disease. The biotechnology epoch of the 20th century that focused on a single gene–single protein approach just doesn't work well with multi-gene disorders such as cancer or Alzheimer's disease. In order to understand the basis for human genetic variability, the human genome project pooled and sequenced the genes and nucleotides of many individuals to establish a baseline. Single nucleotide deviations from this baseline are termed SNPs (single nucleotide polymorphisms). Although rare diseases can occur from SNPs (eg, sickle cell anemia and cystic fibrosis), the most common diseases (eg, diabetes and asthma) may encompass 20–50 SNPs and may involve 10 or more genes. Research efforts are now ongoing to establish blocks of SNPs that correlate with a given disease predisposition. If such correlations can be found, then drugs can be sought to prevent disease expression. The complexity of this undertaking will require a much more sophisticated approach to drug development. Understanding complexity in property-design will also expand possibilities.

Ideally, a property-based design strategy would be able to anticipate and predict the physical properties of a proposed molecule from structure alone. This would be coordinated with activity-based and safety-based strategies so that predictions would be made on this triad of design characteristics. Proposed molecules could then be evaluated from structure alone to see if they either had (a) the requisite properties, or (b) the potential to be further designed to have the requisite triad of design characteristics: activity, solubility, and safety. For this latter group, knowledge of functional groups that have the flexibility for being modified would have to be identified so that further predictions could be carried out on modified structures for triad characteristics. Property-base possibilities would include compounds that had:

a. Passive diffusion properties (solubility & membrane permeability)
b. Crystal packing disruptive potential for passive diffusion

c. Special vehicle delivery potential
d. Prodrug enhancement potential
e. Stability enhancement potential.

FORWARD-FOCUS VISION

Some of the terminologies that we have inherited from crisis situations like attrition and triage cast images of what is to be avoided and what choices have to be made with limited resources. While it is necessary to recognize these areas, a focus on them may inhibit forward thinking and new solutions to get where we want to go. The 'forward focus' model is an alternative way to think about producing more products that add shareholder value. The principles of the model are[12]:

(1) If we focus on obstacles, we expend time and energy on obstacles rather than on getting where we want to go.
(2) When we clearly focus on where we want to go, we do whatever we need to do to get there with minimal wasted energy.

Ironically, empirical evidence suggests that focusing on obstacles may attract what we want to avoid.[12] The forward-focus vision concentrates on the efficient utilization of resources to enable more NCEs to come to market faster, and with higher quality. Its advantage over an attrition-focus strategy is that more energy is expended using existing knowledge to enlarge property-space possibilities and on the development of novel approaches. It has been said that[13] "In the realm of possibility, we gain our knowledge by invention." We also invent rules, but these must be used with caution.

LIFECYCLE OF RULES—Rules are the compilations of knowledge that enable us to carry out business efficiently. Even the best rules, however, should be viewed in the context of a lifecycle. Changing circumstances or new knowledge can cause rules that were formulated in the past to become inappropriate. One of the most useful roles rules play is that they provide a reference for obtaining a more precise understanding of physical phenomena. Attrition also can be thought of in terms of a lifecycle and be made productive.

MAKING ATTRITION NON-PERISHABLE—While late clinical-stage attrition is very costly, the loss of resources involved in attrition of NCEs prior to Phase I clinical trials is even more costly. It is possible that more that 85% of pre-Phase I activity is taken up by compounds that never progress to clinical trials. While this is accepted as an inevitable part of the research and development process, a program for capturing the knowledge from all of these failed NCEs might very well enhance the efficiency of property-design.

ACCEPTANCE OF COMPLEXITY—Rules that capture the essence of complex phenomenon is one strategy for designing properties. Another approach is to accept that physical systems will be complex and that computational approaches may be needed to design systems that can accurately predict. Such systems can analyze more situations in more detail than an individual. One key element that enhances acceptance of such computational approaches is that the reasoning or scientific basis of the predictions be understandable. For continuous progress, phenomena need to be understood at the molecular level.

MOLECULAR PRINCIPLES

Grasping the structure of a subject is understanding it in a way that permits many other things to be related to it meaningfully. To learn structure, in short, is to learn how things are related.[14] Insight that will lead to improved property-based design will result from using a variety of molecular tools that will give scientists an understanding of the precise interactions that occur between molecules, whether they be interactions between molecules among themselves or between molecules with biological systems. The two types of molecular interactions that we will be focusing on in this section deal with interactions in (1) crystals and (2) membranes.

Crystalline interactions are of interest because they ultimately determine solubility, melting point, and dissolution of NCEs. If we can gain a molecular understanding of the intermolecular interactions that occur between the molecules in a crystal, then we can gain insight into how we can predict and design molecules that have the properties we desire from structure alone. This is the ultimate goal of property-based design. For simple crystals, containing only the same molecules (no solvents or salt counterions), we will use the term *cohesive* to characterize the type of intermolecular interactions of the same type of molecule.

Membrane interactions between an NCE and a biological membrane will be termed *adhesive*, because they are between different types of molecules. Adhesive interactions are those types of interactions that also occur between solvent molecules and the NCE when it is dissolved in the aqueous environment of the digestive tract. Solvent-solute interactions control the familiar like-dissolves-like concept. For example, lipid molecules dissolve in oil more than they do in water. We refer the reader to the work of Abraham[15] for extensive research into the solvation phenomena. In this discussion of molecular property-based design, we will begin to examine the types of cohesive interactions that can occur in a crystal which impact its solubility (or insolubility).

Crystalline Interactions

Molecules in a crystal organize themselves in a limited number of regular arrays, which are termed space groups. There are 230 possible crystalline space groups; however, because pharmaceutical molecules are complex and in general not symmetric, the number of actual space groups for drug-like molecules is only about 3. These are shown in Table 38-1. The impact of regular ordering of molecules in a crystal is that, for a given space group, rules can be stated that allow the entire crystal to be replicated through a sequential series of translation, reflection, inversion, and other analytical geometric operations. For example, the operation for the very common space group for drug-like molecules, P21/c, is shown in Figure 38-3. The fundamental unit that is replicated is the unit cell. This is obtained from single crystal x-ray diffraction evaluations of the NCE. This unit cell (sometimes termed the asymmetric unit) has information regarding the number of molecules in the asymmetric unit and the dimension and angles of the unit cell.

Ultimately, it is the molecular structure of the molecule that determines the space group and the number of molecules in the unit cell of a particular crystal. However, for a given molecule, the crystals that can form are not unique. Because molecules can assume different conformations, and because a variety of crystallization conditions can influence the crystal that forms, a variety of different polymorphic forms are possible (this will be discussed in detail in later sections). Polymorphic forms may have different physical properties, especially dissolution characteristics that could impact bioavailability and very often these different forms can interconvert. One objective of active pharmaceutical ingredient (API) design is to find the most stable crystalline form so that polymorphic changes do not occur once an NCE is formulated into a dosage form. It is the packing of the atoms in a given crystal that will be considered next and the forces that lead to insolubility.

CRYSTAL PACKING

Crystal packing is dominated by two opposing phenomena: (1) maximizing the number of hydrogen bonds (H-bonds) that can be formed for a given molecular structure, and (2) packing the atoms of the crystal as densely as possible (ie, close packing). Ultimately, molecular shape and the distribution of the H-bond donor and acceptor groups in a given molecule determine the most favored polymorphic form chosen by nature.

Table 38-1. Possible 3-Dimensional Crystalline Space Groups

CRYSTAL SYSTEM	NUMBER OF INDEPENDENT PARAMETERS	PARAMETERS	MATHEMATICAL ABUNDANCE	ORGANIC CRYSTAL ABUNDANCE
Triclinic	6	$a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$	2	High ?
Monoclinic	4	$a \neq b \neq c$; $\alpha = \gamma$; >90	13	High $P2_1/c$
Orthorhombic	3	$a \neq b \neq c$; $\alpha = \beta = \gamma = 90$	59	Very Low $P2_12_12_1$
Tetragonal	2	$a = b = c$; $\alpha = \beta = \gamma = 90$	68	~ 0
Trigonal rhombohedra	2	$a = b = c$; $\alpha = = \gamma \neq 90$	6	~ 0
Trigonal hexagonal	2	$a = b = c$; $\alpha = = 90$; $\gamma = 120$	19	~ 0
Hexagonal	2	$a = b = c$; $\alpha = = 90$; $\gamma = 120$	27	~ 0
Cubic	1	$a = b = c$; $\alpha = \beta = \gamma = 90$	36	~ 0

H-bonds are non-covalent interactions that can occur within a given molecule (intramolecular) and between different molecules (intermolecular). Essentially they are electrostatic in nature and as such are long-ranging forces (force varies as l/r^2). Weak H-bonds usually have a higher multiplicity of interactions than strong H-bonds because they are more flexible, as illustrated Table 38-2. Intramolecular H-bonds form when the atoms in the molecule can be arranged such that a ring of covalently linked atoms (usually 6) is closed with 1 or more H-bond (Fig 38-4A). Intermolecular H-bonds form between different molecules of a crystal (Fig 38-4B–E).

High affinity traps with their associated insolubility and high melting points can be attributed to H-bonding networks and/or close packing. As a general rule, H-bonding network insolubility is associated with the number of H-bonds per molecule as well as the number of H-bond between molecules in a crystal. In Table 38-3, pairs of molecules are shown that have the same water solubilizing groups but differ in their H-bonding motifs. Figures 38-4B and C show molecules that form a dimer and a single chain, respectively. Each has 2 H-bonds per molecule but differ in the number of H-bonding neighbors. Similarly, Figures 38-4D and E show molecules that form single and double H-bonding chains, respectively. In this case, each molecule has the same number of H-bonding neighbors, but has a different number of H-bonds per molecule. For both pairs, Table 38-3 shows that increasing either the number of H-bonding neighbors or the number of H-bond per molecules reduces the effectiveness of the water-solu-

bilizing group. The negative influence of close packing on physical properties is most likely due to the introduction of van der Waals dispersion forces that vary as l/r^6. Zwitterion formation, conformationally restricted molecules, or high packing density molecules have the highest intrinsic insolubility potential.

Membrane Interactions and Permeability

THEORIES OF PASSIVE PERMEABILITY

The *water of desolvation* hypothesis, explored extensively by Burton and co-workers[16,17] states that the major barrier for passive permeability NCEs across cell membranes is the energy needed to remove bound water from the molecule so it can enter the hydrophobic portion of the lipid bilayer. Although both hydrophobic and hydrophilic NCEs would have some bound water associated with them in solution, the adhesive H-bonding between water and the polar groups of hydrophilic NCEs group would be much stronger and thus need to be broken before transport can take place. Strong supporting evidence for this concept has been found using the peptide bond as the polar moiety and has led to an experimental partitioning system, $P_{heptane/ethylene\ glycol}$, that appears to be more predictive of permeability than the widely accepted octanol/water partition coefficient.[16]

The *molecular rigidity hypothesis* posits that molecular weight itself is not a sufficient condition to impart reduced membrane permeability but may itself be a factor that is correlated with the number of rotatable bonds and polar surface

3-Dimensional unit cell
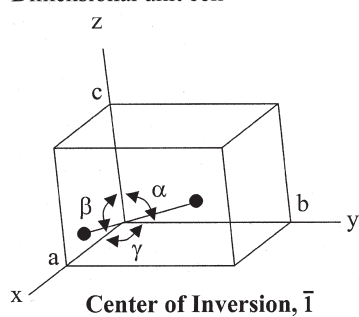
(a) **P ī space group**
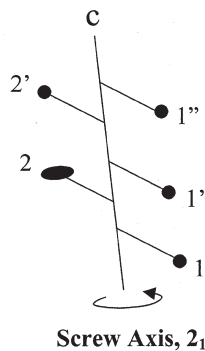
(b) **P 21/c space group**

Figure 38-3. Repeat mechanism (space group rules).

Table 38-2. Comparison of Hydrogen

	BOND CHARACTERISTICS	
	WEAK	STRONG
Bond Character	Electrostatic Broad	Covalent Narrow
Bond Length	1.5 Å–3 Å	1.2 Å–1.5 Å
Directionality	$160° \pm 20°$	$\sim 180°$
Multiplicity	2,3,4 Centered A	2 Centered A
XH ---- A	XH $\vdots$ A′	XH ---- A′ $\vdots$ A″
2 Centered	3 Centered	4 Centered

area. If these two latter parameters are below certain values, then compounds that are sufficiently rigid and non-polar may be absorbed independent of molecular weight.[18] Some factors that can impart rigidity besides fused-ring systems are molecules that have intramolecular H-bonds that form a ring or cyclic peptides.

THEORIES OF ACTIVE PERMEABILITY

NUTRIENT UPTAKE MECHANISMS—The passive permeability limitations discussed above for polar or ionized molecules do not hold for a number of nutrients. Special site-specific transporter proteins are present in membranes that are used to bypass the lipophilic barrier of bilayer membranes.[19] Among these are transporters for peptides, amino acids, nucleoside and nucleobase, ascorbate, and a few other molecules such as glucose and urea. Application of the PEPT1 transporter to prodrug delivery will be discussed below.

XENOBIOTIC EFFLUX MECHANISMS—Membrane transporters belong to one the largest classes of proteins, termed ABC (ATP binding cassette) proteins that can transport against the concentration gradient of the substrate. The characteristics of these membrane proteins are: (a) 2 transmembrane domains [regions of the protein embedded in the membrane], and (b) 2 ABC units [which bind ATP].[20] Defects in ABC proteins are the cause of many human inherited diseases. In most studies, ABC proteins are the multidrug resistance proteins (MDR) that remove therapeutic agents from cells by an active efflux.

MDR1 (or Pgp1) is one of the most extensively studied ABC proteins. Its normal function is believed to protect cells and organisms from toxic substances.[21] There are 7 identified proteins that have been placed in the MDR family, all are organic anion transporters. MDR1, MDR2, and MDR3 have all been associated with multi-drug resistance.

PASSIVE-DIFFUSION DESIGN

One way to reduce conformational restriction is to open up a restricting ring. Alternatively, Figures 38-4 A, D & E discussed in a previous section shows that the substitution of a t-butyl group for a phenyl group dramatically increased solubility by

breaking up H-bonding so that each molecule only had 2 rather than 4 H-bonds per molecule. This was due to the bulkiness of the t-butyl group that prevented dimer formation.

PRODRUG DESIGN

Often NCEs have adequate biological activity but do not have the required physical properties to become a drug. For orally administered drugs, the compound needs to dissolve in the gastrointestinal tract and be absorbed by the intestinal membranes; for intravenous drugs, the compound must have adequate solubility in its dosing vehicle and in the blood so it can be delivered safely without causing embolisms. Prodrugs are one way to solve a number of safety and property-design problems and should be considered early in the design phase. Prodrugs are inactive analogs of biologically active compounds that can be converted into active compounds by the body's chemical processes. They are designed to have the critical properties that the parent compound lacks. Poor membrane permeability, poor solubility, and poor dissolution are problem areas that may be addressed by prodrugs. All three of these areas impact the passive absorption of drugs. Prodrug design has also been used to reduce toxicity.

Poor Membrane Permeability

One of the major roles of the outer limiting membranes of cells is to isolate it from its surroundings. Three factors that inhibit the passage of a drug molecule through biological membranes are: (a) charge, (b) water of hydration, and (c) molecular size. The importance of charge is related not only to the hydrophobic environment of the bilayers but also to the asymmetry of plasma membranes. Because these membranes are composed of two layers of phospholipids (a bilayer), the radius of curvature of micron-sized cells requires that phospholipids with small head groups be located in the inner leaflet of the bilayer to prevent excessive tension on the membrane.[22] The anionic phospholipid, phosphatidylserine (PS), resides almost exclusively in the inner leaflet due to an active process.[23] This negatively charged inner leaflet of the plasma membrane has
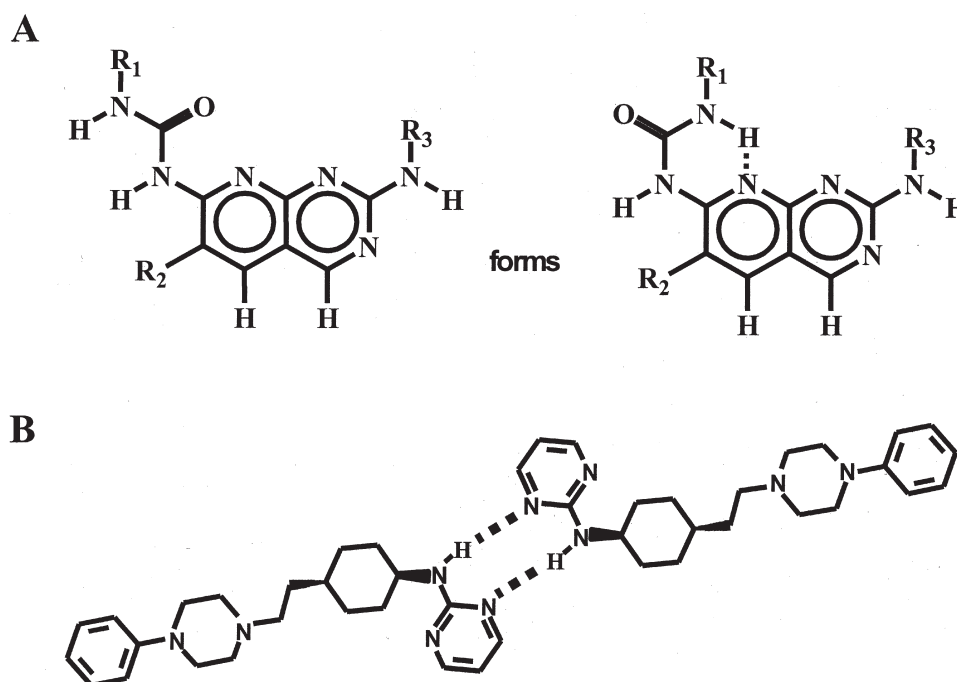
Figure 38-4. Examples of intra- and intermolecular hydrogen bonding.

C

D

E
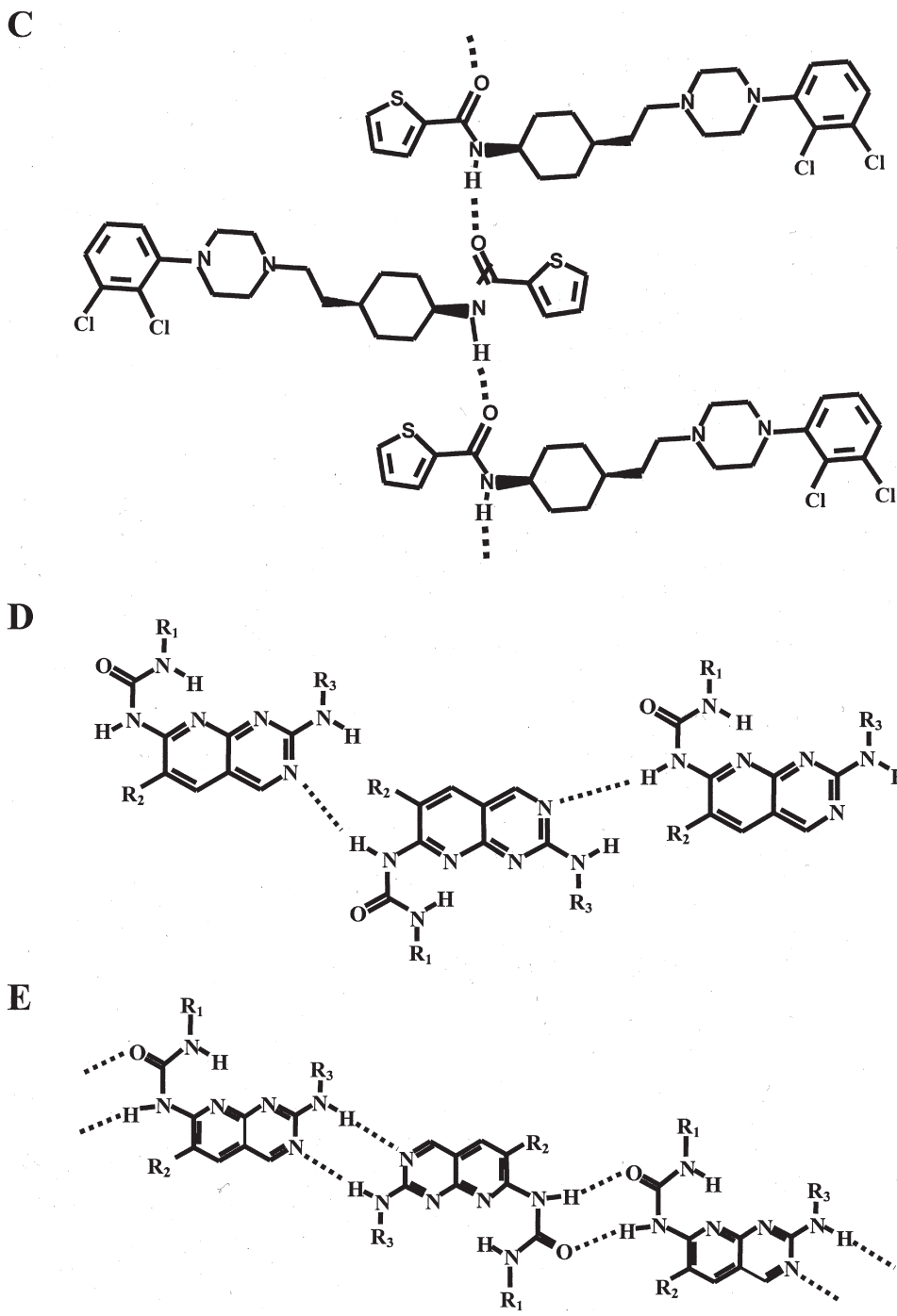
Figure 38-4. *Continued.*

Table 38-3.

pH	SOLUBILITY μg/mL				NETWORK TYPE	# H-BOND /MOLECULE	# H-BONDED NEIGHBORS
	1	5.6	7.3	13			
B	17600		8		Island	2	1
C	14		0.05		Sgl. Chain	2	2
D		1700	610	25	Sgl. Chain	2	2
E		16	10	6	Dbl. Chain	4	2
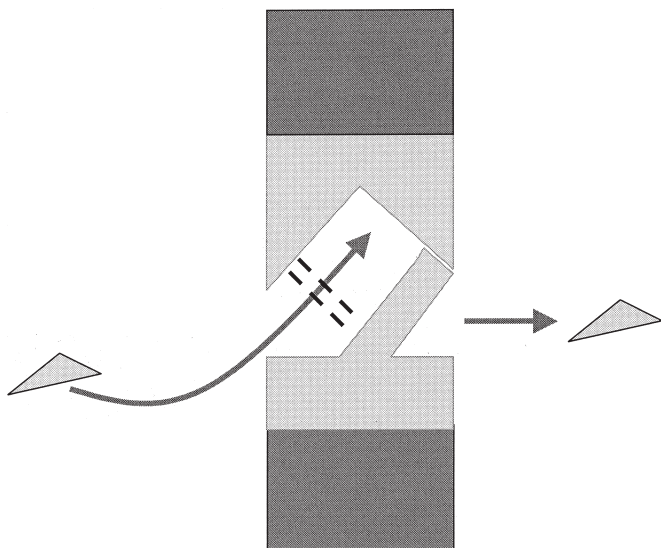
Figure 38-5. PepT1 cattle-gate mechanism.

been shown to control the tissue distribution of basic cationic drugs[24] and the permeability of the anthracycline base, doxorubicin, in a biphasic manner.[25] One might expect this inner leaflet would impact the absorption of anionic drugs. To circumvent these barriers to ionized and polar nutrients like peptides, amino acids, and nucleoside bases, cells developed a number of special transport proteins. Prodrug efforts are now ongoing to exploit these membrane transporters to enhance drug absorption.[26]

Use of Membrane Transporter Systems

Recently, some of the structural requirements of the plasma membrane peptide transporter, PEPT1, have been elucidated.[27-29] The binding requirements and the cattle-gate mechanism for PEPT1 are shown in Figure 38-5. Among the number of drugs reported to be transported by PEPT1 are ACE inhibitors (captopril, enalapril, lisinopril), penicillin, and cephalosporins (ceftibuten, cefadroxil). The advantage of this transporter is its high capacity (grams/meal). Successful prodrug strategies utilizing PEPT1 have been reported. The antiviral agent, Valtrex (valcyclovir-*GlaxoSmithKline*) is a prodrug of Zovirax (acyclovir). It has recently been observed that the H-bonding of the guanidine moiety of L-valaciclovir may enhance its PEPT1absorption.[30]

Reducing Ionization

Most Factor Xa inhibitors for preventing the activation of thrombin and blood clots have utilized a highly charged group, either a guanidine or an amidine group. These groups, however, limit the bioavailability of these compounds when used orally. One strategy to overcome this problem is to synthesize a prodrug which has a reduced charge for oral absorption but which can be converted in the systemic system to the active charged compound. Scientists at *Millennium* have recently designed a Factor Xa inhibitor that utilizes amidoximes as prodrugs for amidines.[31] These prodrugs showed good bioavailability but the conversion to the amidine was only 20%. Although the amidoxime prodrug approach apparently has been successful in masking charge for other chemical entities, in this situation, steric factors evidently retarded activation *in vivo*. This raises another concern with prodrugs: the potential toxicity of the intact prodrug moiety.

The pentamidines are very effective antimicrobial agents against a variety of pathogens and have been used to treat malaria and leishmaniasis. However, their use has been limited to systemic injections since a doubly charged drug is poorly absorbed. Exploration of amidoximes as prodrugs for amidines[32] has led to a new agent, DB 289, that has excellent bioavailability and is currently undergoing phase II clinical trials to treat *Pneumocystis carinii*, a fungal infection in infants that have immune deficiencies and in AIDS patients.[33] Studies with Caco-2 cell monolayers indicate that the greater permeability of the prodrug is due to its ability to transport passively across cell membranes by the transcellular route compared to the pericellular route of the parent compound.

Reducing Water of Hydration

In a previous section, the desolvation hypothesis was discussed in which the impact of strong H-bonds between NCE polar groups and water provides barriers for absorption (due to the need to remove this water before traversing the hydrophobic environment of bilayer acyl chains). Using prodrug strategies to make polar groups more lipophilic is one method to increase permeability and this has been accomplished for peptides by designing cyclic compounds that encourage intramolecular H-bonding and thus reduce water of hydration, make a more compact, rigid molecule, and minimize adhesive interactions with the membrane phospholipid head group.[34]

Size of Molecule

Although molecular weight has always been considered an important determinant of permeability, questions have recently arisen regarding the exact molecular property that determines a reduction of permeability with increasing molecular size as discussed in a previous section. We have discussed the hypothesis that increased molecular rigidity and a reduced polar surface area may enhance permeability. Results with cyclic peptides would seem to be consistent with this hypothesis as the Type I β-turn both reduces the polar surface area and enhances molecular rigidity. In addition, a molecule with more conformational flexibility would appear to present a larger size entity to the membrane.

POOR SOLUBILITY

Using prodrugs for solubility enhancement can take at least two different pathways: (a) increasing water solubility, and (b) disrupting crystal packing. The latter application has as much promise as the first, yet it is less obvious. The reader is referred to the previous discussion on crystal packing. Enhancing ionization with phosphate moieties has been used for both intravenous and oral applications. The intravenous is the earlier.

INCREASING IONIZATION—Fosphenytoin (Cerebyx-*Pfizer*) is an injectable, phosphate prodrug of phenytoin (Dilantin- *Pfizer*) for the treatment of epilepsy that is freely soluble and rapidly cleaved to phenytoin after injection (half-life 8–15 min). The aqueous solubility of the parent drug is $20–25\mu g/ml$ while the solubility of the prodrug is significantly greater (approximately $88,000\mu g/ml$). Local toxicity (pain, burning, itching) that is associated with phenytoin administration due to its high pH formulation is greatly reduced since the more highly soluble prodrug can be formulated at physiological pHs.[35]

DISRUPTING CRYSTAL PACKING—Parecoxib sodium (Pharmacia) is a good example of using prodrugs to disrupt H-bonding and crystal packing as well as increasing pK_a to enhance solubility. For post-surgical pain management, a compound must not only be effective and have few side effects, but it must also be formulated so that a minimal injection volume is administered. Although valdecoxib (*Pharmacia*) possessed
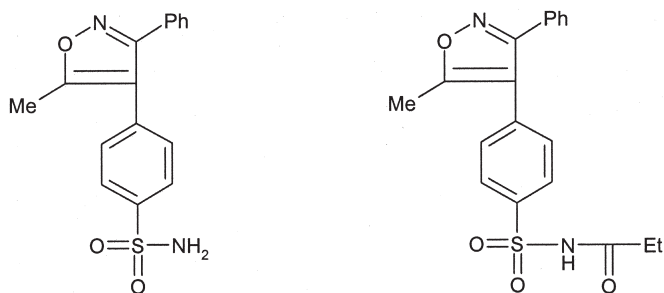
Parent: Valedcoxib
Solubility = 9 µg/mL

Prodrug: Parecoxib
Solubility = 44 µg/mL

Figure 38-6. Prodrug of valdecoxib increases solubility by decreasing H-bonding.

the required potency and safety profile, its solubility was insufficient for this application. Increased water solubility was imparted to the prodrug, parecoxib, by making a prodrug of valdecoxib (Fig 38-6).[36,37]

POOR DISSOLUTION

Prodrugs may be used to improve dissolution properties. For example, Fosamprevavir (Vertex - *GlaxoSmithKline*) is an oral prodrug of Amprenavir (Agenerase - Vertex - *GlaxoSmithKline*), an anti-viral for HIV infections. Although agenerase is approved for HIV treatment, its poor water solubility necessitated that the drug be formulated with large amounts of excipients for optimal dissolution and bioavailability. Typical clinical dosage routines included dosing at 1200 mg (8 capsules) twice or three times a day when plasma concentrations fell below therapeutic levels. The large number of capsules and the food and water restrictions associated with administration of this drug provide barriers to patient adherence with the prescribed therapeutic regimen. By synthesizing the highly soluble phosphate prodrug, fosamprevavir, it is anticipated that adequate drug levels can be achieved with out food or water restriction at 2-700 mg tablets twice daily.[38] Currently, fosamprevavir is completing Phase III clinical trials.

TOXICITY REDUCTION

Xeloda (capecitabine - *Roche*) is a prodrug of the anti-cancer drug 5FU.[39] The parent compound has a number of dose-limiting side effects including: myelo-suppression, intestinal toxicity, and reduction in bone marrow function. Capecitabine reduces the intensity of these side effects by utilizing intestinal, liver and tumor enzymes to generate 5FU in the tumor cell. Camptosar (irinotecan HCl - *Pharmacia*) is a second line agent for advanced colorectal cancer. It is a prodrug of the natural alkaloid camptothecin[40] that is activated by carboxylesterase-2 when it occurs in the tumors. This prodrug greatly increases the solubility of camptothecin.

Taxol's (paclitaxel - *Bristol-Myers Squibb*) low aqueous solubility has necessitated that its intravenous formulation include Cremophor EL which has serious side effects. Recently, a prodrug, paclitaxel oleate, has been shown to not only be activated *in vitro* and in rabbits, but also has been shown to have pharmacokinetic parameters superior to paclitaxel.[41] This raises the possibility of using the most widely prescribed anti-cancer agent with much greater safety. In addition, *Merck* scientists have shown that prostate specific antigen (PSA), a serine protease with chymotrypsin-like activities enzyme, can be used to convert the inactive prodrugs of doxorubicin[42] and vinblactine[43] into the active agent within the tumor thereby reducing side effect of the parent drugs.

In summary, Figure 38-7 shows three types of drug possibility spaces for property-design. The first, at the bottom of the triangle, shows the traditional drug space for compounds that have adequate physical chemical properties and have been found by traditional discovery techniques. The second possibility space is shown in the middle section of the triangle. This space requires more active participation by the property designer to utilize all available tools when physical chemical problems arise. The techniques listed here for simplicity include special delivery systems (SDS) such as self-emulsifying drug delivery systems, prodrugs to break up crystal packing or to add water solubilizing or lipophilic groups, SDS for lipophilic prodrugs, and crystal packing disruptions designed to reduce H-bonding interactions and dense crystal packing. Technology will produce even more options for the future. Finally, there is the physiologically negative drug space or the region of high-affinity traps. These molecules usually have extremely high *in vitro* activity, but have been so over-designed for activity that they suffer from poor physical chemical properties. Sometimes these molecules can be delivered to the systemic system with clever formulations or drug delivery systems, but their poor physical properties ultimately reveal themselves when they crystallize out in the renal tubules of the kidney when solubilizing factors have diffused away from the drug molecules. The ability to anticipate the second possibility space and to avoid the negative-property space at the top of the triangle is a worthy goal for property-based design. This is the subject of the next section.

MACHINE LEARNING SYSTEMS

Artificial intelligence (AI) is a computational algorithm that would be called intelligent if a human exhibited it. One of AI's theses is that computers can simulate any effective procedure.

General Characteristics

- High mp, High crystal energy
- Lipophilic or Less H-bonding
- Less dense packing & Less H-bonding
- Higher Aq. Sol. or Less H-bonding
- Lipophilic or Moderate H-bonding
- General Rule of 5 compliance
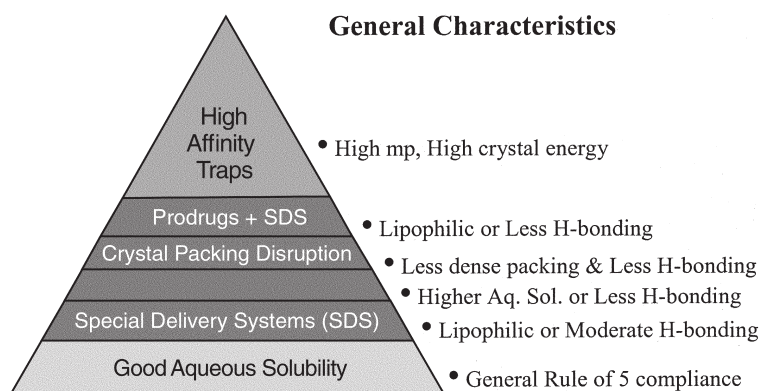
Figure 38-7. Possible and physiological-negative drug spaces. See Color Plate 5.

As John von Neumann once said: "Tell me what a machine cannot do, and I will always be able to make a machine that can do it!" Opponents of AI once defined intelligence as learning. Machine learning is AI's response to that challenge. In the following sections, *machine* will be used synonymously with a computational algorithm.

Machine learning is an area of AI that develops techniques that allow computers to, in some sense, "learn." If the pharmaceutical industry is to become more efficient and reduce cost, it must learn more efficiently. Since 90–95 % of the resources that are expended on NCE development are spent on compounds that will never advance, learning from this experience is an imperative. Machine learning may be the way the industry can reduce cost by learning before doing as we have shown in Figure 38-2. Activity-based design utilizing rapid machine learning techniques would efficiently use the results of high throughput screening to develop highly accurate pharmacophores. In addition, *in silico* activity screening and chemical route design technology would generate structures that are synthesizable, scalable, and match different aspects of these pharmacophores. Safety-based machines would accurately predict different features such as mutagenicity, clastogenicity, or QT-interval prolongation. And finally, property-based machines would be used to ensure that the design of such structures had the requisite physical properties so that traditional or specialized drug delivery could be accomplished. All of these activities would be carried out before a single molecule was synthesized. The impact on cost reduction of such a learning-before-doing paradigm also opens up new markets for NCEs.

Supervised learning is the most prevalent form of machine learning that is currently practiced. Because data in machine learning are termed examples, supervised machine learning is termed *learning by example*. In this type of learning, examples are presented to the machine, and after learning takes place, the machine is tested to see how well it can predict *unseen* examples. Just how accurately the machine can predict unlearned examples is termed the machine's *generalizability*. Example sets are usually subdivided into *training* and *test* sets to carry out the operations stated above. In general, the quality of a machine's future generalizability is highly dependent on how representative the training example set is of examples that are to be predicted in the future.

There are two main types of applications for machine learning, *regression* and *classification*. In regression, the goal is to predict an exact value of a physical property such as solubility or melting point. For classification, the training set is composed of both *positive* and *negative* examples. After training, the machine is asked to correctly separate unseen examples. Classification applications for machine learning are generally *binary* classification i.e. yes/no answers. For example, in the bioinfomatics area, classification is used to predict whether a particular gene codes for a particular protein.

Unsupervised learning deals with learning the *structure* or *topology* of knowledge. Learning that fails to have an ability to grasp the general principles or the structure of a discipline will fall short of learning how things are related and how new information can be related in the future.[44] Learning 'without a teacher' is learning that *adapts* its behavior without being told (supervised learning) the appropriateness (reinforced learning) of an observation. However, by grasping the topology of the subject area, the learning machine will be more able to respond in an improved way in the future. Knowledge discovery and data mining are areas where this type of learning has immediate applications.

One of the major concerns in the machine learning community is the *opaqueness* of some of the algorithms. Humans, and especially physicians, distrust 'black boxes' even if they can be shown to be highly accurate. This concern has lead to new machines that are much more *transparent* in their reasoning. This leads to exciting collaborations between machines and domain experts, humans that are highly specialized in certain technical areas. *Expert systems* are *non-learning* computing systems in which the knowledge of the human domain expert is captured and stored as a set of rules in a knowledge base. A generic inference engine connects the user with the knowledge base so that the machine expert can respond to queries from the user. Machine learning systems, in distinction to expert systems, learn rules from data alone. This is potentially much more powerful since machines can examine data in larger quantities and more consistently than humans. If this process is transparent to humans, it provides a synergistic situation in which the domain expert and the machine can collaborate in solving new problems.

Property-design is based on the premise that all of the information that is needed to predict physical properties is contained in the molecular structure of the molecule alone. This means that the dependent variable (a physical property like solubility) must be computed from factors (independent variables) that are determined from the molecular structure only. The machine learning terminology for these independent variables is *features*; the molecular modeling term for these variables is *molecular descriptors*. There are many computational programs that can generate molecular features and a number of strategies for *feature selection*. The danger, however, is that users get caught up in 'group think' and become so dependent on software programs that innovative thinking is inhibited.

Several mathematical issues are associated with the algorithms of machine learning. The first is the functional relationship of the physical property with features. *Linear* relationships are the simplest type of functional dependence. The advantage of *linear regression analysis* is that humans can easily see and understand the relationships between what is being predicted and the features that are being used to predict (*transparency*). Visual inspection can be used to assess the quality of the prediction. Assuming that there is a linear dependence is both a strength and weakness of this type of analysis. On the one hand, linear system analysis is amenable to many different mathematical analytical methodologies, and, fortunately, many nonlinear systems are linear over a narrow range of feature values. On the other hand, because most physical systems are non-linear over wider ranges, linear dependencies are accurate *locally* but often do not project to the same accuracy over wider ranges (ie, *globally*). Neural networks made the next advance in making predictions. They address the non-linear issue.

Artificial neural networks (ANN) are mathematical abstractions of a simple animal reasoning systems. These systems utilize a non-linear function, usually the hyperbolic tanh function, to model the relationship between the input features with respect to the output physical property. During the learning phase of ANNs, feature selection takes place on the training examples. Learning is a supervised reinforcement that focuses on minimizing error in the training set (empirical risk minimization). The features that have the strongest relationships to the dependent property are selected while taking into account multiple feature interactions. This learning process is often tedious and requires experienced personnel. More over, the complexity of the interactions or the dependence of the dependent property on the input features is hidden, i.e. the reasoning is *opaque*. Another issue with ANNs is that they are subject to *over fitting*. This is a phenomenon in which the ANN model is refined to such a degree that the training examples are very highly correlated to the dependent property but the model as a whole has very poor *generalalizability*. This is a result of learning being dependent on empirical risk minimization. Skilled usage of ANNs, however, can give us some of the most accurate machine learning predictions we have at the current time. In addition, one of the shortcomings of ANNs, a lack of memory, appears to have been addressed. ASNNs were designed with this defect in mind.

Associate neural networks (ASNN) address the issue of training set dependence and knowledge update[45,46] by combining ANN and K-nearest neighbor technology. With such machine learning technology, extensive and laborious training is carried out to generate ensembles of ANNs. The machine has the ability to determine the most appropriate ANN for a

particular compound so that it can obtain the advantage of higher local accuracy while having a global span. In addition, it has the ability to learn new examples on-the-fly. This means that extensive training can be carried out on public databases while updating with respect to proprietary data is possible on an ongoing basis. Recent implementation of an ASNN for calculated LogP has shown 2–5-fold improvements using additional proprietary examples.[47] ASNNs partially address the local/global issue, but still suffer from being opaque. A newer machine learning paradigm has been introduced that addresses both of these issues, *support vector machines* (SVM).

SVMs are statistically constrained machines that were introduced in 1982 to explicitly address *generalizability*, *local/global,* and *linear/non-linear* issues.[48,49] In addition, some SVMs are *very transparent* and are very efficient in *feature selection.*[50] SVMs use mathematical functions, called kernels, that have a very special property: they can act as mediators that allow non-linear data to be processed by linear algorithms. Their major strength is that they promote generalizability explicitly. In addition, SVMs are designed so that they converge on global optima only. They have been shown to give classification results superior to ANNs in the bioinfomatics area and some have regression capabilities. These machines use dual optimization routines that promote generalizability, global, non-linear, and feature efficient predictions, and are just being introduced into the chemoinfomatic arena.[51] In general, however, they are *opaque* techniques that require skill in parameter selection. One machine learning technique, however, excels in its transparency, *inductive logic programming* (ILP).

ACTIVE PHARMACEUTICAL INGREDIENT-BASED DESIGN AND PREFORMULATION

Once a NCE is selected for development, choosing the molecular form that will be the active pharmaceutical ingredient (API) is a critical milestone because all subsequent development will be affected by this decision. For preformulation, physical characterizations should be focused on making decisions that balance solid-state dissolution properties with material consistency under manufacturing and storage conditions. The advantages of having a rapidly dissolving amorphous state have to be balanced against the potential conversion of this state by time, moisture, and heat to a crystalline state that can be less soluble. Similarly, the increased solubility that often can occur with hydrochloride and sodium salts may have to be balanced with a potential for physical or chemical instability due to moisture and heat. These salts are attractive because they are simple to make and are relatively nontoxic. The salt selection process must project its considerations of the "best" properties to encompass dissolution, physical and chemical stability, toxicology, market-image formulations, large scale manufacturing, and product storage.

The following section will outline solid-state changes that might occur with varying moisture content, pH, and temperature. It will be illustrated that water (moisture) is one of the most important environmental factors that influences solid-state stability. The discussion will then focus on identifying the solid-state properties of an NCE that will make it a viable API. Ultimately, the best balance between absorption and material consistency is sought. Later, the discussion of engineering the solid state will explore why these requisite properties should be designed into NCEs from the earliest stages of discovery.

CHALLENGES TO THE SOLID STATE

Solids are a complex state of matter because intermolecular forces can arrange the molecules in a variety of different ways, each producing a different solid with potentially different physical properties. In this section, a symbolic nomenclature is introduced to specifically address changes that can occur in the solid state (Table 38-4). Application of this notation to the ef-

Table 38-4. List of Symbols

SYMBOL	MEANING
α	Amorphous solid state as left subscript designation
Σ	Surface of solid state as right subscript designation
δ	Defective region of solid state as left subscript designation
ρ	Density
I, II, III	Crystalline polymorphic forms of the solid state as left subscript designation
$+$	Positively charged, cationic species as superscript designation
$-$	Negatively charged, anionic species as superscript designation
0	Uncharged, free species as superscript designation
A	Active ingredient in the solid state
a	Dissolved form of the active ingredient
$_j A_\Sigma^{\,i}$	Surface of active ingredient of charge i and solid state j
B	Reactant of A in the solid state
b	Dissolved form of reactant
C_S	Saturation concentration
h	Monohydrate as left subscript designation
$0h$	Anhydrous as left subscript designation
nh	n-Hydrate as left subscript designation
$<h$	Reduced water content as left subscript designation
$>h$	Increased water content as left subscript designation
m	Mass
An^-	Negatively charged anionic counterion
i	Charge on the active ingredient as superscript designation
j	Solid state form of the active ingredient as left subscript designation
k_d	Dissolution rate constant
k_r	Recrystallization rate constant
P	Permeability
Cn^+	Positively charged cationic counterion
S_a	Surface area

fects of moisture, the major environmental factor influencing the solid state, will then be examined.

SOLID-STATE CHARACTER

In this chapter, $_j A_\Sigma^{\,i}$ is a notation that will be used to indicate solid-state changes. The A denotes the active drug entity. This may be a weak acid, a weak base, or a nonelectrolyte. When A dissolves, a denotes the presence of this entity in solution; thus, dissolution of the solid A in water to form a will be shown schematically as

$$A \xrightarrow{\text{H}_2\text{O}} \alpha \tag{1}$$
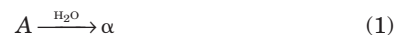
The charge of A is denoted by the usual placement of a right superscript, i. The charge of A is assumed to be zero by default. For emphasis, a lack of charge may be shown explicitly as A^0. For a weak acid, A^0 represents the protonated form (in other notations this might be shown as HA). The ionized form of the weak acid, A^-, represents A^0 minus the weak acid proton. For a weak base, A^0 denotes the uncharged base that can be protonated to $A^0\text{H}^+$. Equations with A, shown with arrows, are not stoichiometric. Instead, they only show essential changes, so the focus can be placed on the relevant chemical, ionic, and solid-state alterations in the chemical entity. For example, in Equation 2, in which a chemical reaction changes the parent entity A into a different molecular solid B,

$$A \rightarrow B \tag{2}$$

there is no attempt to show the specific details of the functional groups that were changed to bring about the formation of B. In a similar manner, consider a reversible acid–base reaction

$$A \underset{\longleftarrow}{\longrightarrow} A^i \tag{3}$$

where i as a plus sign ($+$) represents the cationic form, or a minus sign ($-$) the anionic form, of A. The protonation or deprotonation of a weak basic or acidic group on A will simply be reflected in the charge change that occurs. The scheme is nonstoichiometric because counter ions and charge-balance considerations have not been included.

When a particular molecular organization or emphasis of the solid state is needed, it will be denoted with the left subscript j. A wide variety of different solid states, denoted by $_jA$, are possible. For example, amorphous solids that have randomly packed molecules are denoted as $_\alpha A$ in this chapter. Crystalline solids, on the other hand, have regular packing arrangements and are denoted in a number of ways. Two types of crystalline phases, polymorphs and solvates, are possible for a given molecule depending on the crystallization conditions.

Polymorphs are crystals that have the same molecule formula but have different crystal structures. The Roman numerals I, II, III, . . . are used to denote polymorphs; the most stable polymorph under ambient conditions is usually designated with Roman numeral I. This solid-state form of A will be denoted as $_\text{I} A$ in this chapter.

Solvates, on the other hand, are crystals in which a solvent is incorporated into the crystal structure (polymorphs of solvates could exist). The solvent may be highly bound in the crystal or it may be more loosely bound in channels within the crystal. To simplify this discussion, only water of solvation will be considered. Hydrated solids are denoted by $_{n\text{h}}A$, where n is a fraction or an integer. For example, $_{h/2}A$ denotes a hemihydrate while $_{3h}A$ denotes a trihydrate.

In some situations, it will be useful to emphasize that a particular chemical reaction or physical change is occurring on the surface of a particle. For these purposes, the right subscript Σ will be used to emphasize the surface of the solid state. It should be noted that the right superscript i, used for charge designation, and the left subscript j, used for solid-state designation, are only general placeholders for more specific instances that will be detailed below; on the other hand, the right subscript Σ specifically denotes the surface of a solid particle and not a more general entity. For most situations, the full notation will not be used.

In actual APIs, crystal defective regions A_δ are present. These were formed during large-scale synthesis and milling operations that reduced the API's particle size. In Figure 38-8, defective regions as well as crystalline and amorphous regions are shown diagrammatically.

WATER: A MAJOR ENVIRONMENTAL VARIABLE

The presence or absence of moisture is one of the most important environmental factors that can affect solid-state stability. The surface of an API particle can gain or lose water depending on the relative humidity (RH). Figure 38-8 shows how water vapor can form regions of dissolved drug on the surface of the API par-

ticle. The amorphous region would be expected to dissolve the fastest, and the crystalline region the slowest; that is, the rank order of dissolution would be $A_\alpha > A_\delta > {}_\text{I} A$. In the Figure 38-8 diagram, this is indicated by the font size of the saturated dissolved form of A, a_s, associated with each of these regions. This surface coating results in chemical and physical instability.

Chemical Instability: Water as a Molecular Mobilizer

In general, chemical reactivity is slow in solids because of the spacial separation of different reactive components. For example, if a small amount of an impurity that can act as a catalyst is distributed heterogeneously in an API or a dosage form, the overall rate of reaction is limited because the reaction only occurs in microenvironmental regions. However, in dosage forms, most APIs are usually in contact with moisture-bearing excipients and are stress-tested at elevated temperatures and humidity. The presence of an adsorbed layer of moisture increases the catalytic reactivity of the impurity because water, acting as a molecular mobilizer, can transport different chemical species laterally over the surface of the API.[52] Equation 4 shows a chain of reactions from A to a degradant B,:

$$A \xrightarrow{[\text{H}_2\text{O}]_\text{vapor}} a \xrightarrow{[\text{H}_2\text{O}]_\text{vapor catalytic impurity}} b \xrightarrow{[\text{H}_2\text{O}]_\text{vapor}} B \qquad (4)$$

where b is the solubilized form of B. Moisture also induces solid-state changes in A. (Further discussion of moisture-induced chemical instability will be treated in the section *Hydrate Stability: Importance of the Critical Relative Humidity*.)

Microenvironmental pH: Moisture-Induced Sensitivity of Acid/Bases

Acid–base reactivity in the solid-state change will be enhanced by moisture. Equation 5 shows a moisture-induced change of an anionic salt to its free acid on the surface of a drug particle:

$$A_\Sigma{}^- \xrightarrow{[\text{H}_2\text{O}]_\text{vapor}} A_\Sigma{}^0 \qquad (5)$$

Conversely, Equation 6 shows a moisture-induced surface conversion of a cationic salt into its free base,

$$A_\Sigma{}^+ \xrightarrow{[\text{H}_2\text{O}]_\text{vapor}} A_\Sigma{}^0 \qquad (6)$$

where $A^+ = \text{H}A^+$. Because the amount of solid drug is large compared to the amount of moisture, Equations 5 and 6 have been diagramed as irreversible reactions. Such solid-state changes can alter the physical properties of the API. For example, if particles of the sodium salt of an insoluble acid form a surface coating of the free acid as in Equation 5, the dissolution rate of the surface will be retarded. Testing methods are needed during the salt selection stage to anticipate this type of solid-state change (see under *Salt Selection*).
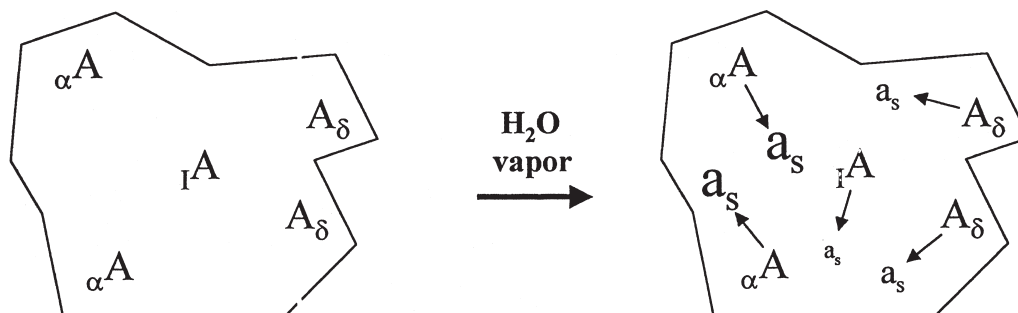

Figure 38-8. Surface of a milled API and dissolution of surface regions due to adsorbed moisture.

Solvent-Mediated Transformations of Polymorphs: Water as a Transporter

If two polymorphic forms can exist at a given temperature, the metastable polymorph will be more soluble (see *Salt Selection*). When this form is put in contact with water, the following solvent-mediated transformation can be promoted:

$$_{II}A \xrightarrow{\text{H}_2\text{O}} {}_1A \tag{7}$$

Water, in the vapor phase, has also been shown to be capable of mediating transformations between amorphous and crystalline forms in both directions.[53]

$$_aA \underset{\longleftarrow}{\xrightarrow{[\text{H}_2\text{O}]_{\text{vapor}}}} {}_1A \tag{8}$$

Finally, transformations can occur that incorporate water into the crystal structure. Here, an anhydrous crystalline form is changed into the monohydrate,

$$_{II}A \xrightarrow{\text{H}_2\text{O}} {}_hA \tag{9}$$

and a salt is transformed into a hemihydrate after passing through the amorphous form:

$$_{II}A^+ \xrightarrow{\text{H}_2\text{O}} {}_aA^+ \xrightarrow{\text{H}_2\text{O}} {}_{h/2}A^+ \tag{10}$$
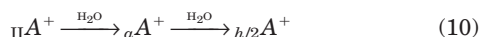
Equations 7 to 10 emphasize solid-state changes. It is likely that most of these transformations may occur only after dissolving and forming a or a species forming a^+.

DECISION-POINTS IN THE DISCOVERY AND DEVELOPMENT OF AN API

The term *active pharmaceutical ingredient* (API), also known as drug substance and bulk pharmaceutical chemical (BPC), highlights both a discovery and a development component. In this section, discovery Steps 1 to 4 will be introduced briefly. The focus will then shift to a detailed discussion of the developmental Steps 5 to 9. Using this background, the section Engineering in the Solid State will outline how early parallel integration of these activities can reduce the time from concept to market.

The term *expansion* is used when choices are being enlarged, and *selection* is used when choices are reduced by decision-making. Ultimately, the expansion and selection phases of discovery lead to a single choice, the best candidate for further development.

1. Library expansion refers to additions to a company's chemical library. Established pharmaceutical companies have amassed hundreds of thousands of compounds through previous discovery efforts. These collections are cataloged carefully and are used systematically in mass screens.
2. Series selection is a decision-making process in which the most active chemicals in the library are identified using a high-throughput biological assay. Typically, these assays are used to detect the ability of a small molecule to interact with a protein, in vitro. In the past, decisions regarding which leads will be pursued further were made based on activity, chemical diversity, patentability, and analog synthetic potential. Today, developmental potential increasingly is part of series selection decision-making.
3. Analog expansion is the increase in the number of compounds targeting a specific activity based on synthetic exploitation of the most promising leads.
4. *Analog selection* is the decision-making process in which the best new chemical entity is chosen for further development. In the past, *in vitro* activity alone was the dominating decision-maker; today, a blend of developmental issues is surfacing earlier.

Preformulation, as well as other areas of development such as metabolism, toxicology, and pharmacokinetics, will play an increasingly important role in Steps 1 to 4. Because a fundamental understanding of the solid state is essential for designing appropriate physical property methodologies for Steps 1 to 4, the remainder of this section will deal with how solid-state proper-
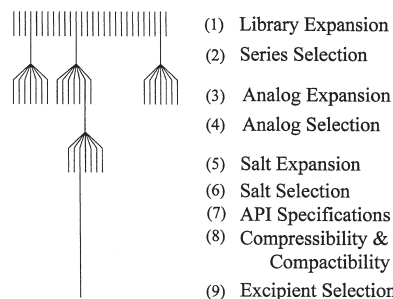

(1) Library Expansion
(2) Series Selection
(3) Analog Expansion
(4) Analog Selection
(5) Salt Expansion
(6) Salt Selection
(7) API Specifications
(8) Compressibility & Compactibility
(9) Excipient Selection

Figure 38-9. Typical API sequential decision-making: selection and expansion cycles.

ties affect absorption and consistency, the two major development issues for an API. Salt selection, which determines the character of $_jA^i$, is the first critical solid-state decision for preformulation in the developmental arena.

Salt Expansion: Exploring the Molecular Possibilities of A^I

The un-ionized (free) form of weak acids and bases, A^0, may not be the ideal molecular form for development. During the salt expansion Step 5 of Figure 38-9, salts are prepared to explore whether one of them would make a more suitable API. Salts are formed by reacting A^0 with an appropriate counter-acid or counter-base. In this discussion, HAn is used to represent a counter-acid that forms an anion An^-. Common counter-acids like HCl and maleic acid are listed in Table 38-5. Similarly, CnOH is used to represent a mineral base of counter cation Cn^+. Common mineral bases like NaOH and KOH are also shown in Table 38-5 along with organic counter-bases.

Table 38-5. Molecular Forms Marketed Worldwide Between 1983 and 1996

SALT FORM	FREQ.	GROUP[a]	PK$_A$	CLOGP	MW
No salt form	390	0			
Hydrobromide	1	1	−8	0.45	80.91
Hydrochloride	102	1	−6.1	0.24	36.46
Sulfate	5	1	−3	−1.58	98.08
Nitrate	6	1	−1.44	2.09	63.01
Phosphate	2	1	2.15	−1.95	96.99
Glucuronate	1	1	3.22[b]	−3.74	194.14
Acetate	8	1	4.76	−0.36	59.05
Maleate	3	2	1.92	−0.18	116.07
Fumarate	8	2	3.02	−0.18	116.07
Tartrate	1	2	3.03	−2.21	150.09
Citrate	1	2	3.13	−2.11	189.10
Succinate	2	2	4.21	−0.62	118.09
Mesylate	8	3	−1.20	−1.31	96.11
Acistrate	1	3	4.91[b]	7.98	284.49
Besylate	2	4	−2.80[b]	0.23	157.17
Tosylate	3	4	−1.34	0.88	171.20
Xinafoate	1	4	2.66[b]	3.00	188.18
Potassium	1	1	16		39.10
Sodium	37	1	14.77		23.00
Tromethamine	2	1	8.07[c]	−3.17	121.14
Bismuth	1	1	1.58		208.98
Bromide	6	5			79.90
Chloride	2	5			35.45

[a] Groups: 0 = No salt, 1 = Polar, 2 = Multifunctional, 3 = Flexible aliphatics, 4 = Planar aromatics, 5 = Quartenary.
[b] Calculated pK$_a$.
[c] Data from *CRC Handbook of Basic Tables for Chemical Analysis,* page 469. From Serajuddin ATM, Sheen P, Augustine MA. To market, to market. In: Bristol J, ed. *Annu Rep Med Chem.* New York: Academic, 1983–1996.

When A^0 is a weak base, the salt, $(A^0H)^+ An^-$, is composed of the protonated form of the base, $(A^0H)^+$ and the ionized form of the counter-acid HAn, An^-. For salt formation, A^0 must be sufficiently basic to remove the proton from HAn (see *Salt-Forming Reactivity Potential*).

Salts have different physical properties than their free forms. Salt selection explores whether a particular salt might have properties that are more appropriate for an API than its parent form. Improving oral absorption by increasing the dissolution rate is often a goal of the salt expansion step. Salts generally dissolve faster in water than their free forms because dissolution is enhanced by the rapid hydration of the ionized salt species with water. Salts of weak bases generally lower the pH of water; salts of weak acids elevate it. For the salt of a weak base in water, the initial dissociation of the salt into the two ions, A^0H^+ and An^- is relatively complete. On the other hand, the deprotonation of A^0H^+ depends on the pK_a of A^0, as shown by these reactions:

$$A^0H^+An^- \xrightarrow{\quad} A^0H^+ + An^- \text{ and } A^0H^+ \underset{\text{high } pK_a}{\overset{\text{low } pH_a}{\rightleftarrows}} A^0 + H^+ \quad (11)$$

It is the release of the H^+ in the second reaction by the salt that lowers the pH and increases the solubility (see *pH-Solubility Profiles*). Hydrochlorides are the most common salts of weak bases.

When A^0 is a weak acid, the salt that forms from a reaction with CnOH is A^-Cn^+ (A^- represents A^0 minus a proton). The most common salts for weak acids are the sodium salts.

Even though salts increase aqueous solubility, they only alter the pH of the solution so that more of the ionized form is present in solution. Salts do not change the ionizable character of the free form; this is an intrinsic property of the free acid or free base and their associated $pK_a(s)$. pH-solubility profiles show the solubility relationship between salts and their free forms.

pH Solubility Profiles

For a weak base, a plot of solubility versus pH will show the highest solubility at low pH and the lowest solubility at high pH; for weak acids, the opposite is true. Such plots give a graphic view of the impact of ionization on solubility for an NCE. The pH range of the small intestine, where oral absorption generally occurs, is approximately 6.5 to 8. It is undesirable to have a compound totally charged or uncharged in this region. If it is entirely charged, there are no un-ionized species that can be transported across the GI membrane. If it is totally uncharged, there are no charged species to enhance solubility. For a monoprotic NCE, the pK_a denotes the pH where the number of charged and uncharged species in solution are equal. On the ionized side of the pK_a, the solubility of the salt limits the maximum solubility. The solubility decline at very low pHs is due to activity and solubility-product effects.[54-56] On the un-ionized side, the solubility of A^0 (the intrinsic solubility) marks the lowest solubility. Salts promote a saturated solution to be formed at a pH that is on the ionized side of the pK_a. They cannot alter the pK_a or the intrinsic solubility. Using these parameters, a qualitative pH-solubility profile can be constructed. Figure 38-10 shows pH-solubility profiles for different counter-acid salts.

The synthesis of salts depends on

1. A proton-exchange reactivity between A^0 and the counter-acid/base
2. A long-range order that permits crystal formation.

The discussion that follows will focus on forming salts from weak bases, because they comprise the majority of the new drug candidates. Weak acids would be treated analogously.

Salt-Forming Reactivity Potential

In order for a salt to form, both the weak base, A^0, and the counter-acid, HAn, must have sufficiently different pK_a values such that a Brönsted-Lowry proton transfer from HAn to A^0 can take place. Table 38-5 gives potential counter-ions and their pK_a values from a listing of all drugs approved worldwide from
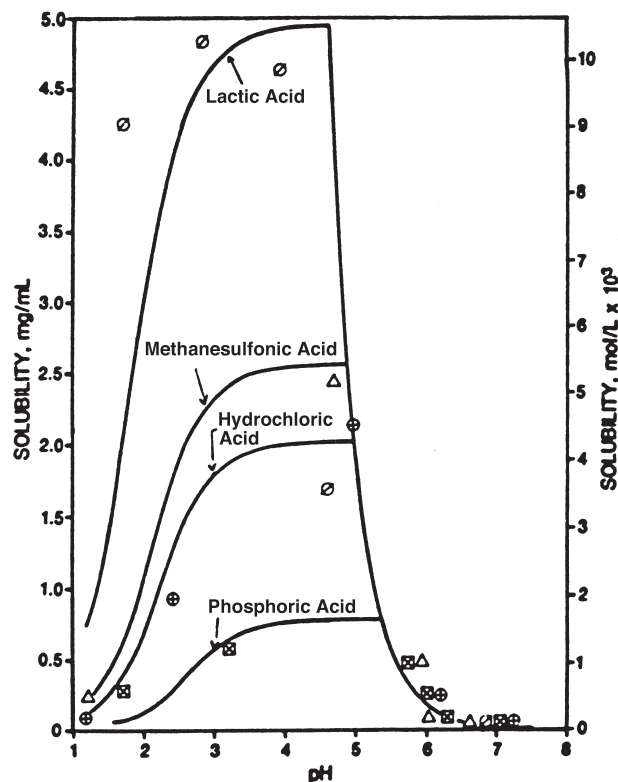
Figure 38-10. pH solubility profile of a weak base. (From Streng WH, et al. *J Pharm Sci* 1984;73:1679.)

1983 to 1996. An acid–base proton transfer should be possible as long as the pK_a of HAn is less than that of the weak base A^0 (recall that the pK_a of A^0 is referenced to its protonated form A^0H^+; see *Solid-State Character*). If ΔpK_a is defined as

$$\Delta pK_a = pK_a \text{ (weak base)} - pK_a \text{ (HAn)} \quad (12)$$

a salt-forming reaction should be possible as long as ΔpK_a is positive. For example, a succinate salt (pK_a 4.2) with doxylamine (pK_a 4.4) is possible[57] where the ΔpK_a is 0.2. Nevertheless, the greater the ΔpK_a, the greater the probability that a salt can be formed. Because the pK_a values in Table 38-5 are calculated for an aqueous environment, this rule must be used only as a guide for salt-forming reactivity in organic solvents. In an organic solvent in which the dielectric constant is lower than water, the ionization equilibria would be shifted:

$$HAn \xrightarrow{\text{low dielectric solvents}} H^+ + An^- \quad (13)$$

$$AH^+ \xrightarrow{\text{low dielectric solvents}} H^+ + A^0 \quad (14)$$
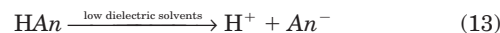
For acridine bases, 50:50 ethanol:water weakens the aqueous pK_a by 1.41 pH units. For the counter-acid, HAn, pK_a weakening is greater than for the protonated base, A^0H^+, because of the greater solubility of HAn in the organic phase and the production of two charges upon ionization. The net effect of organic solvent weakening is to reduce the pK_a difference between the counter-acid and the weak base. This lowers the salt-forming reactive potential. Therefore, in a given organic solvent, if salt formation fails to occur for a particular aqueous ΔpK_a, it is unlikely that salts can be formed in this organic solvent with a smaller aqueous ΔpK_a.

Varying Salt Properties Using Counter-Acid Groupings

For weak bases, salt-forming counter-acids can be used to alter an API's solubility, dissolution, hygroscopicity, stability, and processing.[57] Table 38-5 shows counter-acids organized into dif-

ferent functional groups. For each counter-acid, both the pK_a and the $\log P$ is given where appropriate. A starting point for salt expansion must begin with the properties of A^0. If, for a weak base, $\Delta pK_a = pK_{a\ A}{}^0 - pK_{a\ \text{counter-acid},\ HAn} > 0$, then aqueous salts may be possible. Use of this table and the influence of different counter-acids are covered under *Decision-Tree, Goal-Oriented Approach*.

Crystal Formation Requirements

In general, crystalline solids, including salts, make the most promising APIs. The amorphous form of the solid state is usually not as stable as crystals, either physically or chemically. Crystal formation is a special characteristic of a solid in which the molecules self-organize into regular, repeating, molecular patterns. Solvents play at least three roles in crystallization.

1. They provide some solubilizing capacity so that concentrated solutions can be formed.
2. They promote the nucleation process. Nucleation may be from a pure solution (homogeneous nucleation) or from a seed crystal (heterogeneous nucleation). If a solvent binds too strongly to the molecular organizing functionalities of the salt or seed crystal, crystallization will be impeded. Finding appropriate solvents for crystal formation is a very important step in salt expansion. Failure to adequately explore and find solvents that can crystallize salts could mean that very usable salts would not be evaluated in the salt- selection step because they were not synthesized.
3. Solvents, temperature, and cooling rate can impact the crystal-packing pattern of crystals. Stable polymorphic forms usually are desired for APIs. Metastable forms are normally avoided in an API because they are prone to physical and chemical instability. Solvent conditions that promote metastable and stable crystal formations will be explored under *Metastable Polymorph Formation*.

SALT SELECTION: CHOOSING THE "BEST" API

Salt selection is the first important API decision from the development perspective. Once a salt is chosen, time-consuming and lengthy toxicological studies are initiated that would have to be repeated if the salt form is changed. This decision involves choosing a solid-state phase, $_jA$, which balances potentially conflicting needs: increasing absorption versus maintaining an API that is consistent and can be manufactured in a market-image dosage form (see *Compressibility and Compactibility*). Figure 38-11 shows some of the factors involved in this decision.

Permeability, solubility (C_S), and pK_a are intrinsic properties of A^0 that have been already determined in the analog selection phase (see Fig 38-9). The major dependent variables, absorption and consistency of the API, can be manipulated and balanced in salt selection. In the following sections, the impact of dissolution and particle size on absorption will be explored. In addition, the consistency of the API solid state under the influence of environmental destabilizing factors—such as exposure time (t), ultraviolet light (UV), pH, moisture (H_2O), temperature (T), and pharmaceutical processing operations like milling, compression, and compaction—will be considered.

Absorption Assessment

Oral absorption is generally viewed as two-step, sequential process:

$$A_{\text{solid}} \xrightleftharpoons{\text{dissolution}} a_{\text{GI tract}} \xrightleftharpoons{\text{permeation}} a_{\text{blood}} \qquad (15)$$
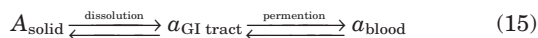
Either dissolution of solid drug, A_{solid}, after the dosage form disintegrates in the GI tract, or the permeation of the dissolved drug, $a_{\text{GI tract}}$, through the GI membrane could be the slowest process. The slower of these two steps determines the overall rate of absorption and is thus rate-limiting.
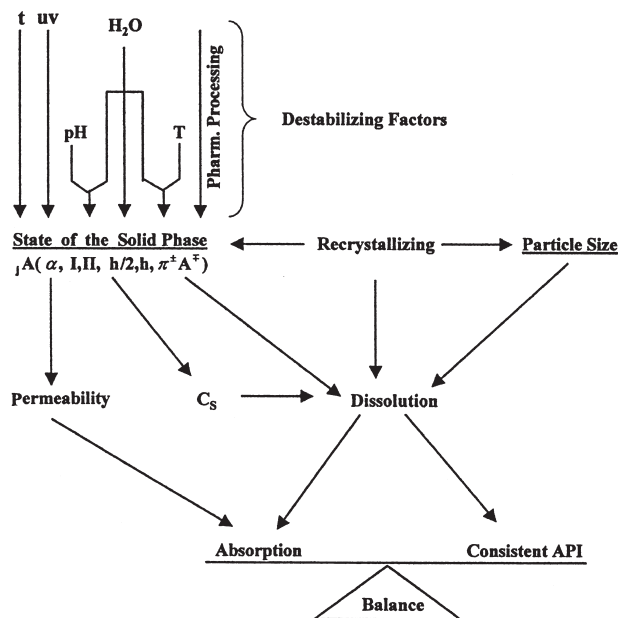
Figure 38-11. API salt selection decision: a balance between absorption and consistency.

Dissolution-limited absorption occurs when the rate of appearance in the GI tract by dissolution (a_{GI}) is slower than the rate of appearance in the systemic system (a_{blood}); *permeation-limited* absorption occurs when the a_{blood} appearance is the slowest process. The impact of these two rate processes on *in vitro–in vivo* (IVIV) correlations will be discussed in the section *Biopharmaceutical Classification of API*. Dissolution-limited absorption will now be considered.

The rate of dissolution of a particle is given by the Noyes–Whitney equation,

$$dA/dt = k_d S_a\,[C_s - C_{\text{bulk}}]\ \text{(non-sink conditions)} \qquad (16)$$

where
A is the amount of drug dissolved.
dA/dt is the rate of dissolution (Q sometimes is used for this rate).
k_d is the intrinsic dissolution constant for the drug.
S_a is the total surface area of the dissolving particle.
C_S is the saturation solubility of the drug at the surface of the particle.
C_{bulk} is the concentration of the drug in the bulk solution.

Because the rate of dissolution depends on the concentration difference between C_S and C_{bulk}, the maximum rate of dissolution would occur if $C_{\text{bulk}} = 0$ (ie, if drug was removed from solution as fast as it dissolved). This would be analogous to a sink that could drain the water coming out of a water faucet as fast as it comes in so that the water level never built up. This analogy is the basis for referring to Equation 16 as nonsink conditions for dissolution, because drug does build up in the solution and the rate of dissolution is correspondingly reduced.

The expression for the maximum dissolution rate is found by setting C_{bulk} equal to 0[58]:

$$dA/dt = k_d S_a C_s\ \text{(sink conditions)} \qquad (17)$$

This initial rate of the Noyes–Whitney equation is termed sink conditions for the dissolution rate.

PARTICLE-SIZE EFFECTS—For a spherical drug particle of radius r, amount m, and of density ρ, Equation 17 can be rewritten as:

$$dA/dt = (3k_d m/\rho)(1/r)C_s \qquad (18)$$

This expression emphasizes the inverse relationship between the dissolution rate, dA/dt, and the particle size r, assuming no dissolution rate-reducing factors are present such as adsorbed air bubbles or aggregated particles.

Smaller particles dissolve faster than larger particles. Thus milling, a pharmaceutical unit-operation, increases dissolution because the API particle size is reduced. On the other hand, when drug particles are suspended in an aqueous solution, particles can increase in size due to recrystallization growth[59]
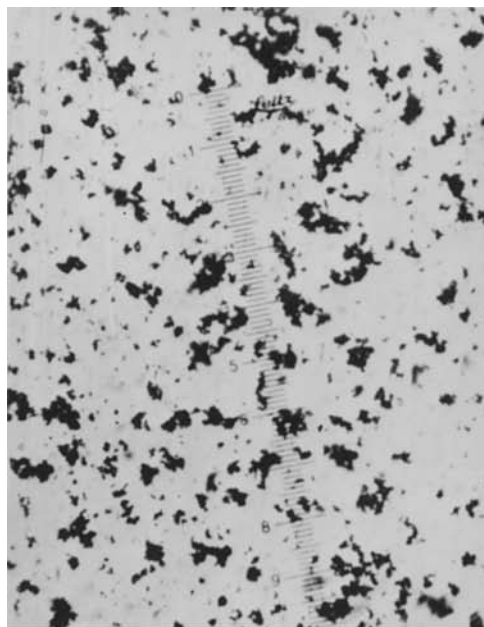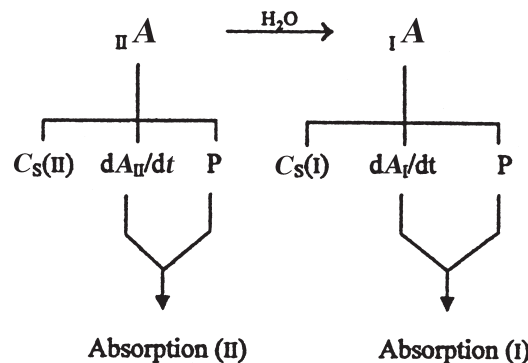

FORM I

INITIAL SUSPENSION

FORM I

SUSPENSION AFTER 6 HOURS.

Figure 38-12. Photomicrographs showing change in crystal size for a suspension of Form 1 of an experimental drug.

Figure 38-13. Absorption changes due to aqueous-phase transformations.

(Fig 38-12). Dosing such suspension orally would be expected to reduce absorption because of a reduction in the dissolution rate.

Reactive Media 1: Implications for Salts of Weak Acids and Weak Bases—When a drug reacts with gastric fluids, its dissolution deviates from Equation 17. For dissolution in 0.1 N HCl, acid–base reactivity is most important for salts of weak acids and for free bases. It has been found that the low pH environment of the stomach dissolves a salt of a weak acid 10 to 100 times faster than the weak acid itself.[60] On the other hand, it is the free base, and not its HCl salt, that dissolves faster in this same environment.[61] These deviations from Equation 17 have been shown to be due to differences between bulk-solution pHs and the pH at the surface of the drug particle. Thus, Equation 17 becomes

$$dA/dt = k_d S_a \, C_{s,h=0} \qquad (19)$$

where $C_{S,h=0}$ is the saturation solubility at the surface of the API.

For weak acid salts, the surface pH has been calculated to be 6.2 to 6.5 for sodium salicylate (pK_a 3.0) and 10.3 for sodium theophylline (pK_a 8.4) in bulk solutions having pHs of 1.10 and 2.1, respectively. On the other hand, the weak base phenazopyridine (pK_a 5.2) sees a surface pH of 3.3 to 3.6, while its HCl salt sees a surface pH of 1.2 for a bulk-solution pH of 1.10. If the solubility due to surface pH and not the pH of the bulk is considered, deviations from Equation 17 become understandable. For the HCl salt, the common-ion effect reduces its solubility from the maximum solubility of the pH-solubility profile at 3.45. Thus, the nonaggregated free base, in this situation, has a surface pH that is optimized to give the highest dissolution rate because it has the highest surface solubility.

Reactive Media 2: Implications for Anhydrates and Metastable Polymorphs—Aqueous-phase transformations are solid-state changes in which water acts as a mediator. During the transition from one form to another, dissolution behavior will reflect the switch from the dissolution rate of the initial solid state to that of the more stable state. Two types of aqueous-phase transformations were introduced in Equations 7 and 9: (1) a transformation from Polymorph II to Polymorph I and (2) a transformation from an anhydrous Form II to a hydrated form h.[62] In Figure 38-13, the transformation of Equation 7 is shown.

Because the permeability (P) of the dissolved drug is the same for the different crystalline forms, the impact on absorption will be due to differences in their solubilities (C_S) as defined in Equation 17 and thus will be reflected in the dissolution rates, dA_I/dt and dA_{II}/dt, being different.

When a solvent-mediated transformation like that shown in Equation 9 occurs, dissolution profiles become more complex. Figure 38-14 shows the biphasic dissolution characteristics for Equation 9. In this situation, an anhydrous substance, $_{0h}A$, becomes hydrated as it dissolves and forms a surface layer of $_hA$. It is this latter layer that controls subsequent dissolution. The concentration versus time plot for the net reaction is $_{0h}A$ (phase change). Note that initially the slope for $_{0h}A$ (phase change) approaches that of the very steep slope $_{0h}A$ (no phase change), and
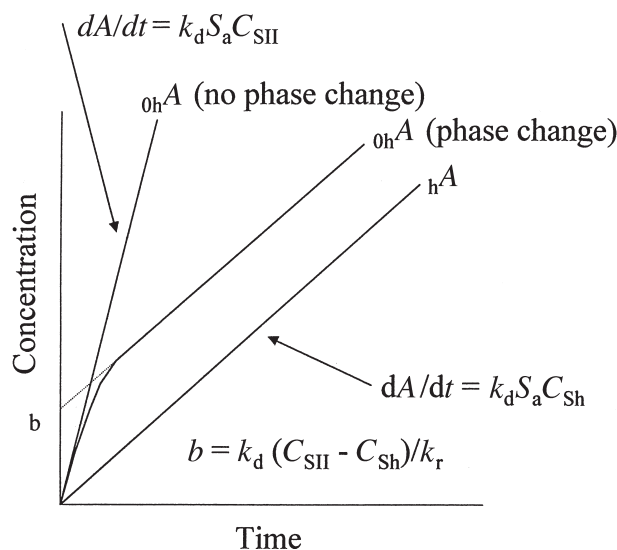
Figure 38-14. Biphasic dissolution of anhydrous to hydrous forms. (Data from Nogami H, Nagai T, Yotsuyanagi T. *Chem Pharm Bull* 1969;17:499.)

that the terminal slope approaches that of $_hA$ (no phase change), the hydrated form. Modifications of Equation 17 to take into account surface recrystallization of $_hA$ on $_{0h}A_\Sigma$ give the biphasic dissolution behavior,

$$dA/dt = k_d S_a \{ C_{sh} e^{-k_r t} + C_{sh}[1 - e^{k_r t}] \quad (20)$$

where k_r is the recrystallization rate constant for the second phase, k_d is the intrinsic dissolution constant, C_{SII} is the saturation concentration for the first phase, and C_{Sh} is the saturation concentration for the second hydrate phase.[63]

## ENHANCED AND RETARDED DISSOLUTION DUE TO SINKS AND PLUGS

The increase in dissolution due to the particle-size reduction of an uncharged API, A^0, can be estimated from Equation 18. Equation 21 shows the resulting surface area increase, $\Sigma\uparrow$, and the corresponding dissolution enhancement.
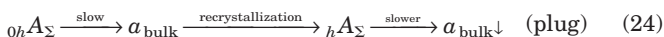
$$A_\Sigma{}^0 \xrightarrow{\text{willing}} A_\Sigma{}^0\uparrow \xrightarrow{\text{faster}} a_S{}^0 \quad (21)$$
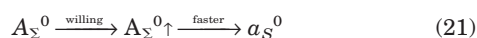
This enhancement, however, is assumed to be under sink conditions and is driven by $C_S = a_S{}^0$ in Equation 17. If the concentration of drug does build up, dissolution is reduced by and is given by Equation 16. This slower dissolution is diagramed in Equation 22 where $a_{bulk}{}^0\uparrow$ indicates the buildup of the drug in the bulk solution.

$$A^0 \xrightarrow{\text{slow}} a_{bulk}{}^0\uparrow \quad (22)$$

An ionizable drug, on the other hand, reduces $a_{bulk}{}^0$, which is indicated by $\downarrow$ in Equation 23 because it is rapidly converted to $a_{bulk}{}^+$, the ionized form. Thus, the ionized form ($a_{bulk}{}^+ = a_{bulk}{}^0H^+$) acts as a sink to remove $a_{bulk}{}^0$ and promotes the dissolution of A^0 by driving the reaction to the right:

$$A^0 \xrightarrow{\text{fast}} a_{bulk}{}^0\downarrow \xrightarrow{\text{very fast}} a_{bulk}{}^+ \quad (\text{sink}) \quad (23)$$
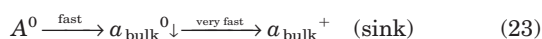
Reduction of dissolution, on the other hand, can occur for an anhydrous API when the hydrated form recrystallizes on the surface as in Figure 38-14. This effect is the opposite of the sink concept, hence the term plugging. Equation 24 show the species involved in plugging. The subscript Σ emphasizes that this is a surface phenomenon.

$$_{0h}A_\Sigma \xrightarrow{\text{slow}} a_{bulk} \xrightarrow{\text{recrystallization}} {}_hA_\Sigma \xrightarrow{\text{slower}} a_{bulk}\downarrow \quad (\text{plug}) \quad (24)$$

ACCEPTANCE CRITERIA GUIDANCE

A simple model to assess the impact of particle size on dissolution and absorption of a non-ionized drug considers the intestine as a single compartment.[63] If the number of particles of uniform size at time t is

$$N(t) = N_0 e^{-Qt/V} \quad (25)$$

where N_0 is the initial number of particles, Q is the flow rate out of the intestine, and V is the intestinal volume, then the surface area for spherical particles of uniform size, r, as a function of time can be given by

$$S_a = 4\pi r^2(t)N(t) \quad (26)$$

This expression can then be used in the non-sink dissolution expression of Equation 16, with certain assumptions including linear intestinal absorption, to approximate the fraction absorbed as

$$F \propto \frac{k_a X_d \hat{t}_r}{X_0} \quad (27)$$

where k_a is the absorption rate constant, X_0 is the administered dose, X_d is the amount of drug dissolved in the GI tract at $\hat{t}_r$, and $\hat{t}_r$ is the GI transit time. Further refinements to this model include accounting for polydispersed spherical powders and comparing cylindrical with spherical shape factors, with and without time-dependent diffusion layer thickness.

Finally, for poorly soluble drugs, simulated dose absorption studies have been carried out over different ranges of solubility, absorption rate constants, doses, and particle sizes. Table 38-6 shows the percent of drug absorbed for a drug that has a solubility of 10 µg/mL with a k_a of 0.01 min^{-1}. Note that, even though particle-size reduction from 100 to 10 µm increases the percent absorbed, as the dose increases, the impact of this reduction decreases dramatically.

Consistency Assessment

POLYMORPHIC STABILITY: IMPORTANCE OF THE TRANSITION POINT

Polymorphic systems, in which different crystalline forms of the same molecular composition can exist, vary in their ability to interconvert at different temperatures. The enantiotropic/monotropic classification is based on the observation that some systems can reversibly interconvert and some cannot. In enantiotropic systems, reversible interconversion between the different forms is possible. For monotropic polymorphic systems, interconversion is only possible in one direction, from a metastable form to a more stable form.

For enantiotropic systems, a critical temperature exists, the transition point, T_p, at which the rate of conversion from one form to another is equal. At temperatures below T_p, one form is more stable; at temperatures above T_p, another form is more stable (see the section *Solid-State Character*; the convention of designating Form I as the most stable polymorph breaks down for such systems because Form I cannot be the most stable form *both* above *and* below T_p).

Figure 38-15 shows a solubility versus temperature diagram for an enantiotropic polymorphic system.[64,65] For the enan-

Table 38-6. Reduced Absorption with Increasing Particle Size for a Poorly Soluble Drug

DOSE	PERCENT OF DOSE ABSORBED			
	25 µm	50 µm	100 µm	
10 µm				
1	91.3	66.9	38.5	17.5
10	70.0	50.0	30.7	15.4
100	9.0	8.7	8.0	6.3
250	3.6	3.6	3.4	3.1

Data from Johnson KC, Swindell AC. *Pharm Res* 1996; 13:1795.
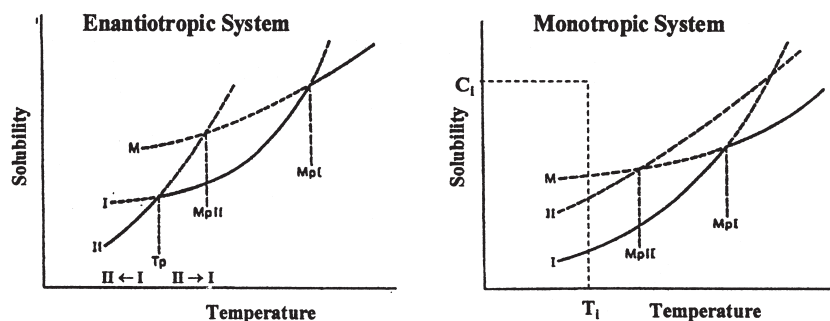
Figure 38-15. Thermal stability of polymorphic systems. (Data from Kuhnert-Bradnstatter M. *Thermomicroscopy in the Analysis of Pharmaceuticals.* New York: Pergamon, 1971; and Heleblian J, McCrone W. *J Pharm Sci* 1969;58:911.)

tiotropic system on the left, at constant pressure, there are three solubility versus temperature curves: Form II is the lowest, Form I is the next higher, and the melting curve is M. The critical temperature, T_p, occurs at the intersection of the Form II and I curves. At this point the solubilities of Form II and Form I are equal and the interconversion rate in any direction is zero.[65] Below the T_p, Form I interconverts to Form II; above the T_p, Form II converts to Form I. The melting point of Form I occurs at the intersection of the Form I curve and the melting curve M.

Because enantiotropic forms show a change in relative physical stability as temperature is changed, it is important to anticipate the impact of temperature on stability. An early warning sign that one is dealing with an enantiotropic system can be found by relating solubilities with thermal parameters. The higher melting Form I has a smaller heat of fusion. Equation 28 gives the relationship between the solubilities,

$$\ln\left[\frac{S_{\mathrm{I}}(T)}{S_{\mathrm{II}}(T)}\right] = \left[\frac{\Delta H_{\mathrm{II}} - \Delta H_{\mathrm{I}}}{RT}\right]\left[\frac{T_m - T}{T_m}\right] \qquad (28)$$

where S_{I} and S_{II} are the solubilities and ΔH_{I} and ΔH_{II} are the heats of fusion of Forms I and II, respectively.[66] The more stable form at a given temperature will have lower solubility at that temperature.

Enantiotropicity exists only when the transition point is below the melting point of Form I (see Fig 38-15). However, if a transition point is not found below the melting point of Form I, it does not mean that the system is monotropic.[65] The transition point, for example, could be below the lowest temperature studied.

For monotropic systems, interconversion is always from the metastable Form II to Form I. The solubility curve of Form II is always above that of Form I, and a transition point does not exist because a crystal cannot be heated above its melting point (see Fig 38-15). Oswald's Law of Stages dictates that if a system is supersaturated with respect to Form II at concentration C_i and T_i, the metastable Phase II will be the first solid phase that appears.[67] As Form II continues to crystallize, the supersaturation is reduced until it reaches its solubility. At this point, although there is no longer a driving force to crystallize more Form II, the solution continues to be supersaturated with respect to Form I. Thus, crystallization of Form I occurs at the expense of the dissolution of Form II.

POLYMORPHIC SOLUBILITY: DIFFERENCE BETWEEN EQUILIBRIUM AND DISSOLUTION-BASED SOLUBILITY

Assume Polymorphs I and II are possible for an NCE. Oswald's Law of Stages tells us that a supersaturated solution will first crystallize out as Form II and then ultimately Form I. Thus, the thermodynamic equilibrium solubility will be limited by the solubility of Form I. However, because the rate of nucleation of II and I is a function of a wide variety of vari-

ables, equilibrium solubility is not an especially useful parameter in estimating the impact of a polymorph form on the absorption of drug from a dosage form. A dissolution-based solubility definition is more useful in this regard. How might such a solubility be defined?

Because the metastable state Form II has a faster dissolution rate, $dA/dt_{\mathrm{II}} > dA/dt_{\mathrm{I}}$, where it is assumed that dissolution is carried out under sink conditions of Equation 17. Because $dA/dt = k_d S_a C_S$, we can conclude that $C_S(\mathrm{II}) > C_S(\mathrm{I})$ if we assume that S_a and k_d are the same for both polymorphs. Thus, Equation 17 provides a working definition for the solubility differences between Polymorph II and Polymorph I, and it provides a method for measuring them from dissolution experiments. More precisely, it provides the solubility at the surface of the API, which is the solubility that is most relevant for dissolution (see the section *Reactive Media 1*).

POLYMORPH CHARACTERIZATION TECHNIQUES

At a given temperature, a fluid-phase transformation can cause a metastable polymorph to change into a more stable, less soluble polymorph. Using a hot-stage microscope, fluid-phase transformations as a function of temperature can be observed.[65] As the temperature is varied, the more soluble polymorph dissolves and the less soluble one grows. If a temperature can be found at which both polymorphs have the same solubility, then the system is enantiotropic, and the temperature is the transition point, T_p. Plots similar to Figure 38-15 can be constructed qualitatively in which the intersection is the measured transition point. These plots are important because they tell which form is most stable at low temperatures, and whether the system is enantiotropic.

Differential scanning calorimetry (DSC) is another characterization tool that is commonly used. It measures heat changes that occur when a solid undergoes phase transitions. Melting of a solid into a fluid, for example, requires an influx of heat into the crystal. Two techniques are useful for detecting polymorphic systems using DSC: scanning-rate variation and temperature cycling.

Scanning-rate variation has been shown to detect some reversible polymorphic systems. In Figure 38-16, crystallization of the more stable polymorph shows up as exothermic depressions as the scanning-rate increases.[68] Hot-stage microscopy can be used to confirm these thermal changes.

Temperature cycling using DSC also can be used to study the relative interconvertability of crystalline forms. A loss of the metastable, lower melting point polymorph of metoclopramide base was found after heating, cooling, and then reheating.[69] The more stable polymorph can often be observed as exotherms due to crystallization after heat–cool cycles.[70] In addition, storage of a metastable polymorph below the melting point of either polymorph can result in the formation of the more stable polymorph. For gepirone hydrochloride, this occurred after a heat treatment of 3 hours at 150° C.[68]
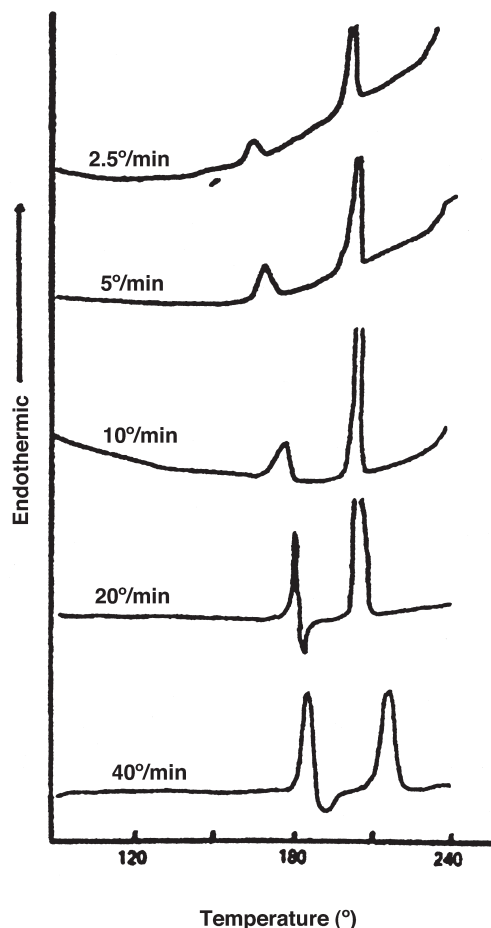
Figure 38-16. Detection of polymorphs by varying the DSC scanning rate.

Powder x-ray diffraction is the most powerful method for detecting polymorphs. Because different polymorphs have different crystal structures, the packing patterns of their atoms are different. Powder x-ray diffraction detects these packing differences as differences in diffraction patterns. Comparisons of diffraction scans between different polymorphs show characteristic differences that can be used for identification (fingerprinting) purposes.

Single-crystal x-ray diffraction is the most definitive characterization tool because the exact relative locations of atoms in the molecular crystal can be determined. However, most often, high-quality crystals for this type of analysis are not available from the bulk API (especially if the material was milled). Recrystallization of suitable crystals from saturated solutions may be possible. If the single-crystal x-ray diffraction problem can be solved, programs are now available that can convert single-crystal diffraction data to a powder x-ray diffraction pattern. This is necessary to ensure that the recrystallization process has not grown a new polymorph.

Solid-state nuclear magnetic resonance (NMR) is also a powerful technique for studying polymorphic systems. In this technique, a powder sample must be rotated at a special angle (the *magic angle*) with respect to the magnetic field so that preferential orientations of the powder particles are averaged. Microcalorimetry also has been used to characterize the thermodynamic properties of different polymorphs. Finally, diffuse reflectance infrared Fourier-transform spectroscopy recently has been used to quantify binary mixtures of polymorphs using the partial least-squares method for spectral analysis.[71]

METASTABLE POLYMORPH FORMATION

Exploring the potential that a given salt has for polymorph formation is a very important aspect of salt selection. It is important that the choice of the final molecular form be based on as much information as possible. Other factors being equal, a molecular entity that forms polymorphs is generally not as desirable as one that does not, because of the potential interconversion of polymorphs and a change in an API's dissolution. This could cause consistency problems both in the API and in the dosage forms. Special techniques are used to attempt to synthesize metastable polymorphs. Preparation of metastable polymorphs requires:

1. Supersaturating conditions for the metastable form, $_{II}A$.
2. Crystallization of the metastable state before the stable polymorph forms.
3. Stable conditions for the metastable polymorph so that conversion to the stable $_I A$ form is prevented.

These steps are shown in Figure 38-17.

For a monotropic system, the metastable state can only change to the stable state; for an enantiotropic system, the transition point is critical for interconversion. Therefore, the formation temperature should be as far above the transition point as practical.

The ideal solution conditions to prevent $_{II}A$ from converting to $_I A$ are such that the solution phase, a, should be highly supersaturated, of a small volume, and in a relatively poor solvent. Rapid cooling is the method of choice for maintaining supersaturation with respect to $_{II}A$. To help ensure that the rate of metastable crystallization is much greater than the rate of thermodynamic equilibration, small volumes and poor solvents for $_I A$ are used. The use of dry ice for rapid cooling with alcohol or acetone is common for these purposes. Once crystallization from the saturated solution phase, a, has occurred, it is important to filter and dry the precipitate as quickly as possible to prevent a fluid-phase transformation to the stable polymorph. Alternatively, if $_I A$ can be melted without degradation, complete melting and rapid cooling of the melt is an another method of forming metastable forms. This avoids two major problems of solution-phase metastable polymorph formation—filtration and drying, both of which can promote interconversion.

HYDRATE STABILITY: IMPORTANCE OF THE CRITICAL RELATIVE HUMIDITY

Relative humidity (RH) is the percentage of the maximum amount of moisture that air can hold. A substance is hygroscopic when it takes up this moisture from air. For a drug substance, the RH that is in equilibrium with a saturated aqueous solution of a solute is termed the critical relative humidity (CRH).[72] It is a key parameter that can influence the physical stability of solid-state hydrates. A number of studies have shown that the gain or loss of water from a hydrate can center on the CRH. Because water in organic crystals is never a passive entity (see *Hy-*
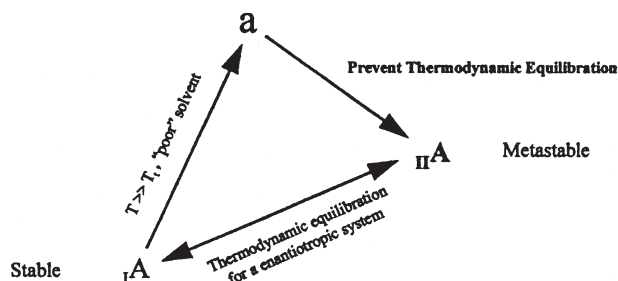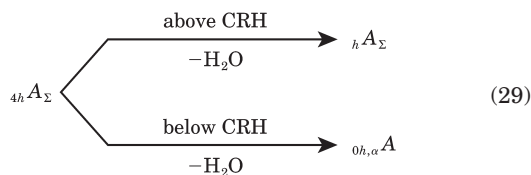

Figure 38-17. Formation of a metastable polymorph in a monotropic system.

drate Formation), solid-state changes in the crystal are very likely to follow.

For the tetrahydrate sodium salt of a tetrazolate derivative, a number of different solid-state forms are possible.[73]

$$
{}_{4h}A_\Sigma \begin{cases} \xrightarrow[-\text{H}_2\text{O}]{\text{above CRH}} {}_hA_\Sigma \\[2em] \xrightarrow[-\text{H}_2\text{O}]{\text{below CRH}} {}_{0h,\alpha}A \end{cases} \tag{29}
$$

The conversion of ${}_{4h}A$ to ${}_hA$ requires elevated temperature and a RH above the CRH. Water's plasticizing action in reducing the intermolecular H-bonding between adjacent molecules is believed to be the mechanism that facilitates the solid-state transformation to the more stable ${}_hA$ crystal form.[74] Similarly, elevation of both temperature and RH were required to convert the ${}_{0h}A$ form of paroxetine HCl to the ${}_{0.5h}A$ form.[75] Water also promoted a solid-state transformation of the αA form to the ${}_{0h}A$ form of a disodium leukotriene antagonist. The amorphous form initially picked up a small amount of water (2%) and then slowly released this water as the anhydrous form was formed. Conversely, the humidity-mediated conversion from ${}_{II}A$ to αA has been observed for another leukotriene antagonist.[76] Difficult hydrate situations have been dealt with by carefully defining the RH ranges of different species and setting specifications consistent with typical manufacturing environments.[77]
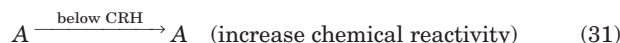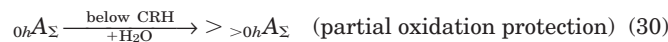
In general, hydrates that are more closely packed tend to be more physically stable with respect to moisture loss. The ideal solid state is one that is stable over a wide range of RH, such as the ${}_{0.5h}A$ form of paroxetine HCl.[75] For the sodium salt of the tetrazole derivative shown in Equations 29 and 30, the denser ${}_hA$ structure is physically more stable than the ${}_{4h}A$ structure. The latter loses four water molecules from crystal channels at a significantly lower temperature than the one water molecule of the ${}_hA$ form, which is integrated into the crystal structure in a more cohesive manner.[73] In the sections *H-Bonding Networks*, and *Hydrate Formation*, hydrate formation is discussed from a molecular point of view. Crystal formation involves two mutually opposing principles: (1) satisfying the molecule's intermolecular H-bonding needs and (2) packing the atoms in the crystal as closely as possible. Hemi- ($h/2$) and monohydrates (h) evidently satisfy both close packing and H-bonding needs more efficiently than hydrates that contain water in channels.

Hysteresis is a general term that is used when a material's response to a second exposure of a stress differs from a prior response. This has been observed in the moisture uptake of an API as a function of RH. A number of instruments are now available that can monitor a sample's weight as RH is cycled from 0% to 95%. The noncoincidence of the weight as the sample is back cycled from 95% to 0% indicates hysteresis. One explanation of this type of behavior is that surface-initiated changes occurred in the solid state below or above the sample's CRH. Dehydration of the surface below the CRH, as in Equation 29, with the formation of an amorphous coat of ${}_{0h,\alpha}A_\Sigma$ means that any subsequent water vapor will encounter a more hygroscopic surface than ${}_{4h}A_\Sigma$ and thus a different hydration kinetic behavior. On the other hand, conversion of ${}_{4h}A$ to ${}_hA$ above the CRH, as in Equation 30, will produce a different kinetic behavior upon rehydration. Thus, RH hysteresis may result from changes in both the kinetic and equilibrium behavior of the surface of the particle.
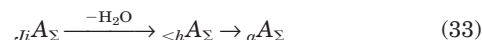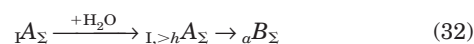
CHEMICAL STABILITY: COMMON DEGRADATION SEQUENCES—BELOW CRH

SORPTION/DESORPTION OF SURFACE WATER—If an anhydrous form of A is exposed to an RH below the CRH, water molecules will slowly adsorb onto the surface of the drug particle (denoted as $>0h$). Adsorption of up to a monolayer of water has been shown to provide partial protection from oxidation. Dehydrated foods, for example, are more stable when moisture coats reactive sites. For the anhydrous phenylbutazone, the oxidation rate has been shown to be lower below the CRH.[78] For a hydrate, however, the loss of surface water of hydration (denoted as h) at RHs below the CRH has been shown to increase reactivity. Equations 30 and 31 show both of these possibilities.

$$
{}_{0h}A_\Sigma \xrightarrow[+\text{H}_2\text{O}]{\text{below CRH}} > {}_{>0h}A_\Sigma \quad \text{(partial oxidation protection)} \tag{30}
$$

$$
A \xrightarrow{\text{below CRH}} A \quad \text{(increase chemical reactivity)} \tag{31}
$$

FORMATION OF AN AMORPHOUS (A) SURFACE—A water enriched/depleted surface, ($>h/<h$), is prone to further solid-state changes shown in Equations 32 and 33. For the water-enriched surface, a chemical reaction is shown in which the crystalline form of A ($j = I$) reacts to form the product αB_Σ, which is amorphous. This type of surface hydrolysis at RHs below the CRH was shown to occur for meclofenoxate HCl decomposition[79] and for propantheline bromide hydrolysis.[80] For the latter, a lag time occurred that was attributed to the amount of time that was necessary to form a monolayer. For the water-depleted hydrate ($j = h$), the loss of water initiated the formation of an amorphous surface layer, αA_Σ. The consequences of these amorphous surfaces will now be explored.

$$
{}_IA_\Sigma \xrightarrow{+\text{H}_2\text{O}} {}_{I,>h}A_\Sigma \rightarrow {}_aB_\Sigma \tag{32}
$$

$$
{}_{Ji}A_\Sigma \xrightarrow{-\text{H}_2\text{O}} {}_{<h}A_\Sigma \rightarrow {}_aA_\Sigma \tag{33}
$$

TRANSFORMATION OF AMORPHOUS SURFACES—Because amorphous layers are more prone to be hygroscopic than crystalline solids, the chemical transformation of ${}_IA_\Sigma$ to αB_Σ in Equation 32 is significant because the latter can attract more water to the surface. Dissolution of αB_Σ shown in the first downward reaction of Equation 34 will then form a surface coated with b_Σ, as shown in Figure 38-8. The reaction of meclofenoxate HCl below the CRH to form amorphous dimethylaminoethanol HCl (see Eq 32) is a good example of this.[79] Next, the water adsorbed to the surface due to the dissolved form of B on the surface, b_Σ, promotes the dissolution of the surface of A, A_Σ, to form a surface coated also with a_Σ, the dissolved form of A on the surface, which then undergoes further decomposition to b_Σ. This is shown in the horizontal and final downward reactions of Equation 34.

$$
\begin{array}{c} {}_aB_\Sigma \\ \downarrow +\text{H}_2\text{O} \\ A_\Sigma \xrightarrow{b_t} a_\Sigma + b_\Sigma \\ \downarrow \\ b_\Sigma \end{array} \tag{34}
$$
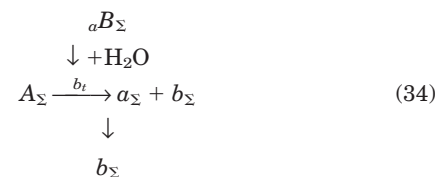
In Equation 35, two possible solid-state changes for αA_Σ are shown. First, the reactive amorphous surface can undergo a degradation reaction to form C_Σ. Second, the surface can continue to lose water below the CRH so that the subsurface ${}_hA$ undergoes a solid phase transformation to a crystalline phase, ${}_IA$. The dehydration changes for cefixime trihydrate are examples of these reactions.[81] The partially dehydrated form of this compound was more unstable than the fully hydrated or the completely dehydrated crystalline forms.

$$
{}_aA_\Sigma \begin{cases} \xrightarrow{T} C_\Sigma \\[1.5em] \xrightarrow{-\text{H}_2\text{O},\,T} {}_IA_\Sigma \end{cases} \tag{35}
$$
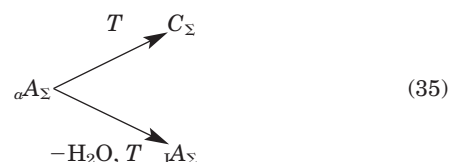
CHEMICAL STABILITY: COMMON DEGRADATION SEQUENCES—*ABOVE CRH*

When water is adsorbed to the surface of the particle above the CRH, the drug particle becomes coated with a dissolved drug layer, a_Σ, which is assumed to be saturated[52]:

$$A_\Sigma \xrightarrow{\text{excess } H_2O} a_\Sigma \tag{36}$$

Degradation under these conditions is assumed to occur solely in the dissolved layer. This situation has been extensively discussed.[52] For the Maillard reaction, in which primary amines react with carbohydrates, adsorbed water initially increases the reaction rate to a maximum due to the enhancement of reactant mobility. Greater amounts of water then decrease the reaction rate due to dilution of the reactive species. Similarly, for free-radical auto-oxidation of unsaturated groups, reactivity increases above the CRH because of accelerated reactant mobility. Below the CRH, oxidation decreases due to the immobilization of hydrogen peroxides and trace metal catalysts and the protective effects of a monolayer of water that is insufficient to increase reactant mobility.

INFLUENCE OF SALT FORM ON HYGROSCOPIC-ITY—Table 38-2 shows that the non-salt forms, including free bases, free acids, and nonelectrolytes, are the most popular molecular forms on the market. In general, these forms would be expected to be less hygroscopic than salt forms due to their un-ionized character. Although the sodium salt is the most popular weak acid form, this form has a tendency to be hygroscopic. Alternative salts that have proven useful in overcoming hygroscopicity are hydrogen sulfate[82] and tromethamine.[83,84]

Hygroscopic tendencies for weak bases might be overcome by using aromatic counter-ions. Aryl sulfonic acids were shown to provide moisture protection without decreasing dissolution for the sparingly soluble weak base, Xiobam.[85] The free-base form of this drug (pK_a 6.1) was hydrolyzed at 40°C/80% RH. On the other hand, one weak base (pK_a 3.67) was chosen for development because it was less reactive to moisture exposure than the HCl salt. The latter showed chemical instability with moisture and heat and was the only salt that could be formed.[86] Stronger bases like pelrinone (pK_a 4.71) can form stable and nonhygroscopic HCl salts.[87]

GRINDING IMPACT—Processing of solids can have a major impact on dissolution due to solid–solid phase changes. Grinding is one process that has been shown to cause changes in both polymorphs and hydrates. For the $_{III}A$ polymorph (Form C) of chloramphenicol palmitate,[88]

$$_{III}A \xrightarrow{\text{grinding}} _{II}A \xrightarrow{\text{more grinding}} _{I}A \tag{37}$$
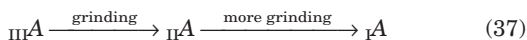
grinding causes a successive change to the $_{II}A$ polymorph (Form B) and finally to the $_{I}A$ polymorph (Form A).[89] Correspondingly, dissolution from the fastest to the slowest is in the order

$$_{\text{ground } II}A > _{\text{ground } I}A > _{II}A > _{I}A \tag{38}$$

For hydrates, similar solid-state changes have been observed. When cefixime trihydrate is ground, a solid-phase transformation takes place:

$$_{3h}A \xrightarrow{\text{grinding}} _{\alpha,0h}A \tag{39}$$

Water in this situation plays an essential role in crystal formation. Its removal causes a collapse of the crystal lattice.[90] Other pharmaceutical processing operations and their impact on crystals have been reviewed. [91]

SALT SELECTION DECISION-MAKING

The pressure to increase the productivity of the knowledge worker is readily apparent at the salt-selection stage. Because of increased productivity in discovery, the cascading impact on development to choose rapidly the best molecular form is readily apparent; toxicological and bioavailability studies cannot proceed until the salt is chosen. Once these studies are initiated, it becomes very costly to change the molecular form because many of these biological studies would have to be repeated. More importantly, precious time and a competitive advantage will be lost. However, if an unanticipated, unacceptable property emerges during the development of an API, the sooner the change is made the better. It is for these reasons that efficient paradigms are being sought for this stage of development. Two approaches will be presented that attempt to optimize the probability of success with speed. Previous approaches were criticized for excessive characterization of poor candidates and for a lack of clear go/no-go decision-making.[92] As a practical consideration, it is essential that NCEs have high purity, and that salts be crystallized. In the following discussion, weak bases that are to be absorbed orally are used. Similar approaches can be developed for intravenous NCEs and for weak acids.

Multi-Tiered Selection Approach

One approach in which different critical parameters are used to filter a salt candidate's progression to the next stage has recently been proposed.[92] Crystalline salts are successively sorted by a three-tier system in the following way:

Tier 1. Hygroscopicity
Tier 2. Thermal analysis and x-ray diffraction
Tier 3. Accelerated solid-state stability

Tier 1 eliminates any form with excessive moisture sorption/desorption characteristics. Only the survivors progress to Tier 2. In this second tier, changes in crystal structure are examined under extremes of moisture conditions by using thermal analysis and powder x-ray diffraction to detect desolvation and aqueous-phase transformation problems. In addition, aqueous solubility is determined to address potential dissolution problems. The best candidates for formulation and manufacturing are considered here and survivors proceed onto Tier 3. In this third tier, accelerated thermal and photo-stability testing is carried out. This is considered to be the most time-consuming step so the limiting of candidates saves time and effort. Selected excipient compatibility testing may also occur at this stage. If Tier 2 eliminates all of the candidates, additional salts or free acid/bases are considered before reevaluating any salt that was dropped in an earlier tier.

Several comments can be made regarding this approach.

1. The HCl salt of ranitidine, due to its hygroscopicity,[93] probably would not have been a final candidate in the multi-tiered approach. Yet this is one of the most successful drugs ever marketed. This emphasizes a need for prioritizing the salt selection process so that as wide of a range of development issues are addressed as early as possible and that they all are put in perspective. If a hydrochloride salt has much better absorption properties than the free base but is hygroscopic, it would be very prudent for development to see if it can deal with this problem. Otherwise, bioavailability may be compromised by a single-minded emphasis on API consistency.
2. The free base is not considered in the multi-tiered approach unless all alternatives have failed despite its potentially favorable dissolution in gastric fluids and its sensitivity to particle size reduction with a reactive sink.

The decision-tree, goal-oriented approach discussed below addresses some of these issues.

Decision-Tree, Goal-Oriented Approach

An alternative approach to the multi-tiered go/no-go selection approach is one based on a decision-tree using statistical probabilities and functional grouping of counter-ions to seek prioritized physical properties. In Figure 38-18, prioritized problems are shown, absorption being the highest priority.
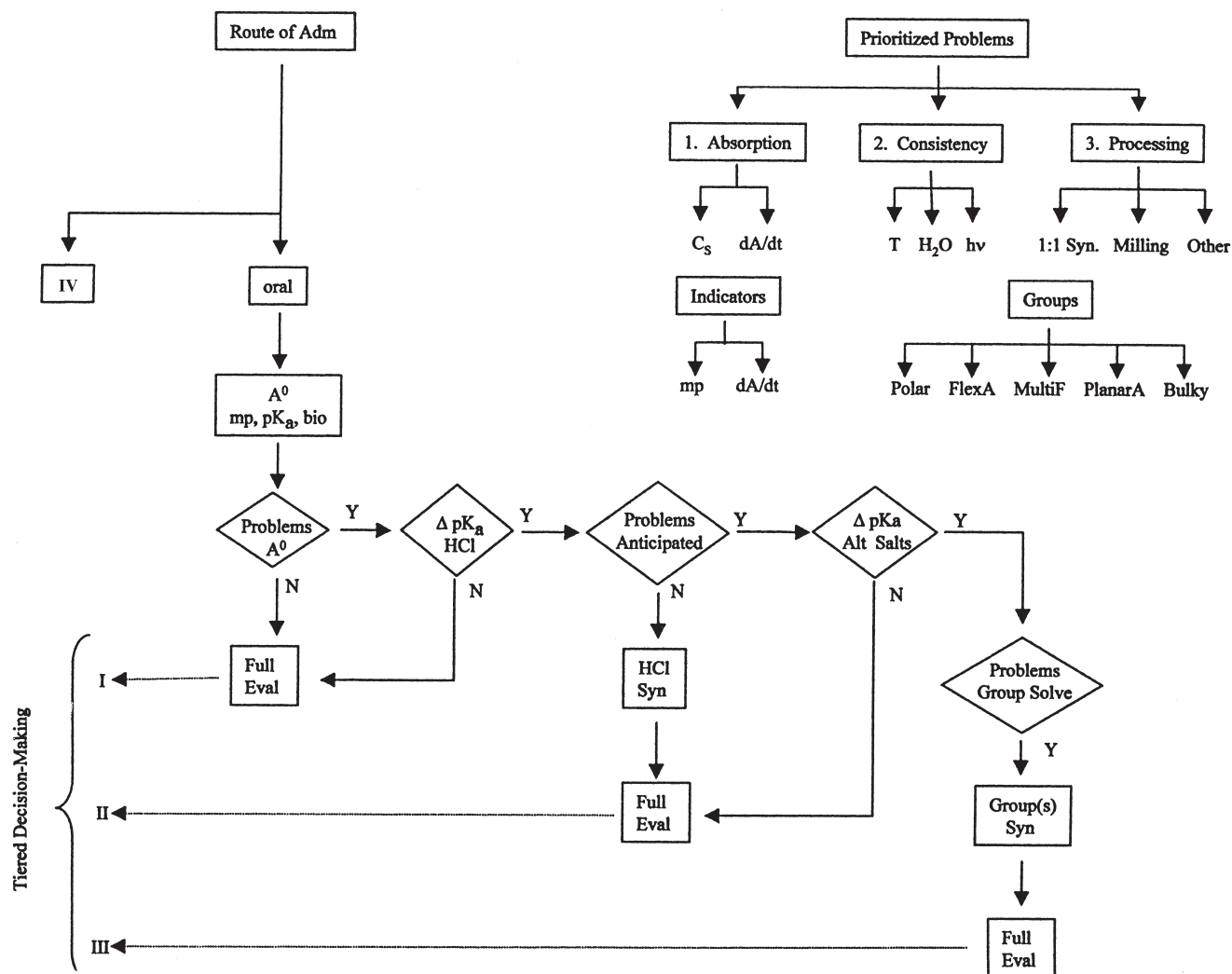
Figure 38-18. Absorption-dominated decision-tree.

The decision-tree considers the free base, the HCl salt, as well as other options. Although this approach uses statistical probabilities for molecular form consideration, ideally, a high-throughput, automated methodology would be available that could determine exhaustively which salts can form crystals and under which conditions. Feasible salts would then be synthesized and placed under accelerated stability and stressing conditions. This would allow for the maximum amount of exposure to the sample before a decision has to be made. Degradant evaluation need not be carried out on these stressed samples immediately; other issues may eliminate a particular candidate and make this unnecessary. However, evaluation for crystallinity should be carried out early to ensure that this does not impact physical or chemical stability. Physical property screens and absorption-dominated prioritization would then force a pharmaceutical evaluation to be made regarding the possibility of overcoming consistency and processing problems.[94] By using functional groupings (see Table 38-5), salt forms would be considered that could address specific problems.[57]

EXCIPIENT SELECTION: FORMULATION COMPATIBILITIES

Excipients serve many roles and are the backbone of a formulation. They may be needed to stabilize the API by providing an-

tioxidant, heavy-metal chelating, or light-protection properties. They also may be used to enhance bioavailability and to control the release from dosage forms. For solid dosage forms, they provide suitable properties for dispensing the API in accurate dosage units that have reproducible release properties. Diluents provide a flowable bulk, binders hold powders together, lubricants provide punch-releasing properties, and disintegrants help to disperse dosage forms in the GI tract. On the other hand, judicious choices must be made to prevent incompatibilities between the API and excipients.

Screens to detect drug-excipient incompatibilities recently have been developed using elevated temperature and added water to accelerate potential interactions in ternary and more complex powder blends.[95] Such methods have been shown to be capable of rapidly detecting chemical incompatibilities and giving good correlations with results using powder blends of drug and excipients at elevated temperatures and humidity.

Processing incompatibilities can be more difficult to troubleshoot than chemical incompatibilities. For example, tablet performance has been shown to vary for ketorolac tromethamine, depending upon the kind of starch that was used. Cornstarch showed a decreased disintegration time and dissolution rate as a function of blending time whereas pregelatinized starch showed no such dependency. The difference between these two excipients was attributed to the formation of drug/cornstarch agglomerates with magnesium stearate.[96]

Blending studies have shown the potential benefits of using sodium lauryl sulfate to offset these types of effects.[97]

Finally, manufacturing for a global market has forced a reevaluation of excipients that are used in formulations so that manufacturing can be carried out with internationally acceptable components. The European Economic Community has recently focused the pharmaceutical industry on eliminating excipients that have the potential for transmissible spongiform encephalopathies, replacing ingredients like stearic acid, magnesium stearate, polysorbate 80, and simethicone with vegetable grade sources.

API SPECIFICATIONS: MEETING PRODUCT AND REGULATORY REQUIREMENTS

Polymorphic Forms and Hydrates Decision Trees

A major portion of this chapter has been devoted to characterizing the solid state, $_jA$. The left side of Figure 38-19[98,99] summarizes some of the potential solid states that can exist for the unionized form of A; if a salt form was chosen for the API, the same states also would be possible. Previous sections have discussed the impact on API consistency and dissolution for the different solid states. The critical relative humidity (CRH) and the transition point (T_p) for enantiotropic polymorphic systems are especially important intrinsic physical parameters that control solid-state consistency and potential solid-state interconversion. Moisture and temperature, as we have discussed, are the major environmental variables that can promote these changes. Rapid methods, therefore, are needed to characterize potential solid-state forms and their physical properties. The decision-tree on the right side of Figure 38-19 summarizes when specifications need to be set to maintain API consistency. If the physical properties of the solid states differ, assessments need to determine the impact this will have on a formulated API. Specifications need to be set to ensure a consistent product.

Particle-Size Acceptance Criterion

Once the solid state, $_jA$, has been characterized, the potential impact of particle size on absorption can be assessed. Figure 38-20 shows a decision-tree approach, suggested by the Interna-

tional Committee on Harmonization, for determining whether a particle-size acceptance criterion is needed.[100] Previous sections in this chapter have discussed nearly every aspect of this tree. Although dissolution-limited absorption is a major concern, Figure 38-20 also includes dosage form issues such as content uniformity.

Biopharmaceutical Classificaion of API

Although it is possible to alter the solid state, $_jA$, such that dissolution and absorption can be enhanced, solubility and passive permeability are, in general, intrinsic properties of the NCE. Thus, even though the amorphous state, αA, in some situations can be stabilized to enhance dissolution, the equilibrium solubility will be determined by the least soluble solid state. A classification has been proposed to segregate situations when *in vitro* and *in vivo* correlations (IVIV) are expected. Such designations may be used as a guide for determining when bioequivalent studies may need to be carried out. Table 38-7 shows the four major classes based on solubility and passive permeability.

CONCLUSION: APPLICATION OF KNOWLEDGE

"The actual product of the pharmaceutical industry is knowledge; pills and prescriptions ointments are no more than packaging for knowledge."[101] The introduction of methods to probe and exploit human and animal genomics has had a cascading impact on the industry. These new concepts had a number of qualities that ensured adaptation. The systematic use of mechanism-based reagents was a tangibly better solution for finding new therapeutic entities than the more serendipitous methods of the past. Such high-throughput screens were compatible with increasing use of robotics whose advantages could easily be understood by all in the pharmaceutical industry. Each company was able to hold trial runs to test the utility of such screens and in the end obtain observable results. Today, the recombinant DNA innovations of the 1980s still provide the driving force for other innovations in the pharmaceutical industry: miniaturization, customizing, and artificial intelligence.

Miniaturization began in earnest with the micronization of the transistor concept onto silicone chips. In the pharmaceutical industry, mass screening, the demand for higher and higher
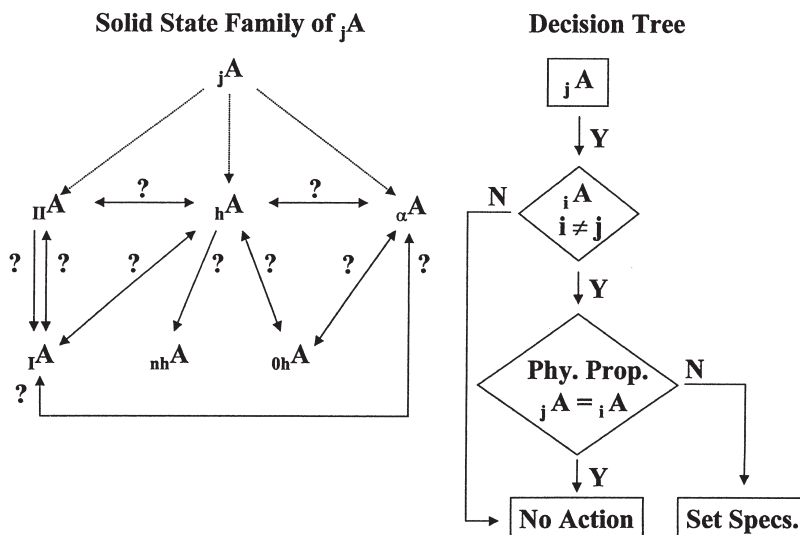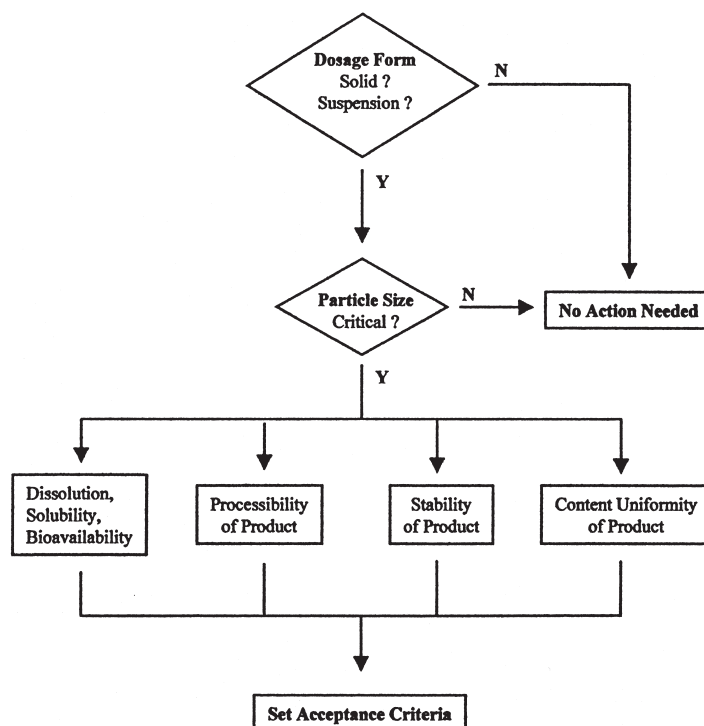

Figure 38-19. Solid-state forms and specification setting. (Data from Byrn S et al. *Pharm Res* 1995;9:84; and Byrn S et al. *Gold Sheet* 1996;30(6):1.)

Figure 38-20. Decision-tree for drug substance particle-size distribution. (From Byrn S et al. Specifications for new drug substances and products: Chemical Substances, ICH4, Fourth International Conference on Harmonization. Brussels, July 1997.)

throughput, and the need to conserve chemical libraries have accelerated analytical and synthetic nanotechnology. This latter need is extremely important because chemical libraries are expendable resources that are not easily replaced. Old library entries were synthesized in gram quantities, and newer entries in milligrams. Conservation of this resource will require a combination of nanotechnology along with a host of regeneration technologies including combinatorial synthesis, high-throughput purification, and promotion of an increasingly diverse molecular library for mass screening. In addition, chromatographic columns, HPLCs, and electrophoresis on the nanoscale hold promise for extremely high resolution with extremely low material consumption. On this scale, area can efficiently be converted to a linear dimension. Thus a chip 10×10 mm can be converted easily to an electrophoretic path of 9.5 cm. The potential for massive parallel processing is evident when one con-

templates the possibilities of 100 nanolaboratories on a single chip.

Customization at low cost also will be possible with new technology. DNA probes located on biochips will permit the individualization of a treatment course depending on a person's ability to metabolize a given drug. Such innovations likely will cause a cascading demand on development to individualize dosage forms. Finally, the rapid and parallel demands placed on preformulation will force more decisions to be made using artificial intelligence. High-throughput determinations of physical properties will result in high quality databases, which can in turn be systematically exploited by expert systems. Highly accurate predictions of solubility, permeability, and dissolution will be possible in the 21st century.

Although artificial intelligence is still in its infancy, the benefits of its applications can be appreciated from a consideration of the differences between knowledge and information. A chemical reaction database, for example, stores information on particular reactions. However, it cannot apply this information to new molecules. Expert systems, on the other hand, so codify knowledge that they can be applied to entirely new situations. Knowledge differs from information in that information is random and miscellaneous, and it tends to expand too rapidly and overwhelm us. Knowledge, on the other hand, requires that the structure of a subject be understood in a way that permits other things to be related to it in a meaningful way; it permits intuitive heuristic procedures to be developed to solve problems when no algorithms are available. Such applications of artificial intelligence, however, are still in the early-stage knowledge revolution, in which knowledge is applied to produce results. In the postcapitalist society, knowledge will be applied toward systematic innovation: "It will be applied systematically and purposefully to define what new knowledge is needed, whether it is feasible, and what has to be done to make knowledge more effective."

Knowledge and the productive application of knowledge are anticipated to be the sole factors that will drive the postcapitalist society into the 21st century. In the pharmaceutical industry, massive diffusion of innovations from discovery into de-

Table 38-7. *In Vitro/In Vivo* Correlation Expectations for Immediate-Release Products Based on Biopharmaceutics Class for Passive Absorption

CLASS	SOLUBILITY	PERMEABILITY	IVIV CORRELATION EXPECTATION
I	High	High	IVIV correlation if dissolution rate is slower than gastric emptying rate. Otherwise limited or no correlation.
II	Low	High	IVIV correlation expected if *in vitro* dissolution rate is similar to *in vivo* dissolution rate (unless dose is very high).
III	High	Low	Absorption (permeability) is rate-determining and limited or no IVIV correlation with dissolution rate.
IV	Low	Low	Limited or no IVIV correlation expected.

From Amidon GL, et al. *Pharm Res* 1995; 12:413.

velopment will pose an accelerating challenge for preformulation. To meet this challenge, preformulation, through a better understanding of the solid state, must seek to design improved characteristics into APIs at the earliest stages of discovery. This will be the edge that any company will need to facilitate the rapid movement of new therapeutics entries to marketplace. The patient is waiting!

REFERENCES

1. Handy C. *The Age of Paradox*. Cambridge, MA: Harvard Business School, 1995: 49–67.
2. Stella V. One of the many fine terms Dr. Stella has coined along with 'grease balls' and 'brick dust'.
3. Lipinski CA. *Adv Drug Del Rev* 1997; 23:3.
4. DiMasi JA. *Clin Pharmacol Ther* 2001; 69:286.
5. Arlington S, et al. *Pharma 2010: The Threshold of Innovation.* IBM Future Series, GS10-9439-00, 2000.
6. Kulling SE, Metzler M. *Food Chem Toxicol* 1996, 35:605.
7. DiMasi JA. *Clin Pharmacol Ther* 2001; 69:297.
8. Pisano GP. *The Development Factory–Unlocking the Potential of Process Innovation*. Cambridge, MA: Harvard Business School, 1997.
9. Sinko PJ. *Curr Opin Drug DiscDev* 1999; 2:42.
10. Caldwell GW. *Curr Opin Drug Disc Dev* 2000; 3:30.
11. Vankatesh S, Lipper RA. *J Pharm Sci* 2000; 89:145.
12. Oakley E, Krug D. *Enlightened Leadership.* New York: Simon & Schuster, 1994:76–93.
13. Zander RS, Zander B. *The Art of Possibilities.* Cambridge, MA: Harvard Business School, 2000.
14. Bruner JS. *The Process of Education*. Cambridge, MA: Harvard University Press, 1960:7.
15. Abraham MH, Le J. *J Pharm Sci* 1999; 88:868.
16. Goodwin JT, et al. *J Med Chem* 2001; 44:3721.
17. Goodwin JT, et al. *J Peptide Res* 1999; 53:355.
18. Veber DF, et al. *J Med Chem* 2002; 45:2615.
19. Anand BS, Dey S, Mitra AK. *Exp Opin Biol Ther* 2002; 2:607.
20. Klein I, Sarkadi B, Váradi A. *Biochem Biophy Acta* 1999; 1461:237.
21. Borst P, et al. *Biochem Biophy Acta* 1999; 1461:347
22. Sheetz MP, Singer MJ. *Proc Nat Acad Sci* 1974; 71:4457.
23. Boon JM, Smith BD. *Med Res Rev* 2002; 22:251.
24. Yata N, et al. *Pharm Res* 1990; 7:1019.
25. Speelmans G, et al. *Biochem* 1994; 33:13761.
26. Anand BS, Dey S, Mitra AK. *Exp Opin Biol Ther* 2002; 2:607.
27. Bailey PD, et al. *Angew Chem Int Ed* 2000; 39:505.
28. Terada T, et al. *Pflugers Arch* 2000; 440:679.
29. Swaan PW, et al. *Receptor Channels* 1998; 6: 189.
30. Friedrichsen GM, et al. *Eur J Pharm Sci* 2002; 16:1.
31. Song Y, et al. *Bioorg Med Chem Lett* 2003; 13:297.
32. Hall JE, et al. *Antimicrob Agents Chemother* 1998; 42:666.
33. Zhou L, et al. *Pharm Res* 2002; 19:1689.
34. Gangwar S, et al. *Pharm Res* 1996; 13:1657.
35. Stella VJ. *Adv Drug Del Rev* 1996; 19:311.
36. Talley JJ, et al. *J Med Chem* 2000; 43:1661.
37. Talley JJ, et al. *J Med Chem* 2000; 43:775.
38. Corbett AH, Kashuba ADM. *Curr Opin Invest Drugs* 2002; 3:384.
39. Shimma N, et al. *Bioorg Med Chem* 2000; 8:1697.
40. Xu G, et al. *Clin Cancer Res* 2002; 8:2605.
41. Lundberg BB, et al. *J Controlled Release* 2003; 86:93.
42. Garsky VM, et al. *J Med Chem* 2002; 45:4706.
43. Brady SF, et al. *J Med Chem* 2001; 44:4216.
44. Bruner JS. *The Process of Education.* Cambridge, MA: Harvard University Press, Cambridge, 1960:7,17–32.
45. Tetko IV. *J Chem Inf Comput Sci* 2002; 42:717.
46. Tetko IV. *Neur Proc Lett* 2002; 16:187.
47. Tetko IV, Tanchuk VY. *J Chem Inf Comput Sci* 2002; 42:1136.
48. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines.* Cambridge UK: Cambridge University Press, 2000.
49. Schölkopf B, Smola AJ. *Learning with Kernels.* Cambridge, MA: MIT Press, 2002.
50. Gunn SR, Kandola JS. *Machine Learning* 2002; 48:137.
51. Burbidge R, et al. *Comput Chem* 2001. 26:5.
52. Shalaev EY, Zografi G. *J Pharm Sci* 1996; 85:1137.
53. Sokoloski TD, et al. *Pharm Res* 1994; 11:S1.
52. Duddu SP, et al. *Pharm Res* 1994; 11: S1.
53. Vadas EB, et al. *Pharm Res* 1994; 8:148.
54. Streng WH, et al. *J Pharm Sci* 1984; 73:1679.
55. Serajuddin ATM, Mufson D. *Pharm Res* 1985; 2:65.
56. Serajuddin ATM, Sheen P, Augustine MA. *J Pharm Pharmacol* 1986; 39:587.
57. Wells JI. *Pharmaceutical Preformulation: The Physicochemical Properties of Drug Substances.* New York: Wiley, 1988:38.
58. Hussain A. *J Pharm Sci* 1972; 61:811.
59. Nielsen AE. *Croatica Chemica Acta* 1987; 60:531.
60. Serajuddin ATM, Jarowski CI. *J Pharm Sci* 1985; 74:148.
61. Serajuddin ATM, Jarowski CI. *J Pharm Sci* 1985; 74:142.
62. Nogami H, Nagai T, Yotsuyanagi T. *Chem Pharm Bull* 1969; 17:499.
63. Dressman JB, Fleisher D. *J Pharm Sci* 1986; 75:109.
64. Kuhnert-Bradnstatter M. *Thermomicroscopy in the Analysis of Pharmaceuticals.* New York: Pergamon, 1971:35–36.
65. Heleblian J, McCrone W. *J Pharm Sci* 1969; 58:911.
66. Rocco WL, Swanson JR. *Int J Pharm* 1995; 117:231.
67. Cardew PT, Davey R. *Proc Roy Soc Lond A* 1985; 398:415.
68. Behme RJ, et al. *J Pharm Sci* 1985; 74:1041.
69. Mitchell AG. *J Pharm Pharmacol* 1985; 37:601.
70. Shah AC, Britten NJ. *J Pharm Pharmacol* 1987; 39:736.
71. Hartauer KJ, Miller ES, Guillory JK. *Int J Pharm* 1992; 85:163.
72. Admirat P, Grenier JC. *J Rech Atmos* 1975; 9:97.
73. Kitamura S, et al. *Pharm Res* 1992; 9:138.
74. Tada T, et al. *J Pharm Sci* 1987; 76:S302.
75. Buxton PC, Lynch IR, Roe JM. *Int J Pharm* 1988; 42:135.
76. Vadas EB, Toma P, Zografi G. *Pharm Res* 1991; 8:148.
77. Morris KR, et al. *Int J Pharm* 1994; 108:195.
78. Yoshioka S, Shibazaki T, Uchiyama M. *J Pharmacobiodyn* 1986; 9:S6.
79. Yoshioka S, Shibazaki T, Ejima A. *Chem Pharm Bull* 1982; 30:3734.
80. Yoshioka S, Uchiyama M. *J Pharm Sci* 1986; 75:92.
81. Kitamura S, et al. *Int J Pharm* 1990; 59:217.
82. Gu L, et al. *Drug Devel Ind Pharm* 1987; 13:437.
83. Gu L, Strickley RG. *Pharm Res* 1987; 4:255.
84. Roseman TJ, Yalkowsky SH. *J Pharm Sci* 1973; 62:1680.
85. Walkling WD, et al. *Drug Dev Ind Pharm* 1983; 9:809.
86. Serajuddin ATM, et al. *J Pharm Sci* 1986; 75:492.
87. Hajdu J, Adams G, Lee H. *J Pharm Sci* 1988; 77:921.
88. Aguiar AJ. *J Pharm Sci* 1969; 58:963.
89. Kaneniwa N, Otsuka M. *Chem Pharm Bull* 1985; 33:1660.
90. Kitamura S, et al. *Int J Pharm* 1989; 56:125.
91. Grant DJW, York P. *Int J Pharm* 1986; 30:161.
92. Morris KR, et al. *Int J Pharm* 1994; 105:209.
93. Teraoka R, Otsuka M, Matsuda Y. *J Pharm Sci* 1993; 82:601.
94. Gould PL. *Int J Pharm* 1986; 33:201.
95. Serajuddin ATM, et al. *Pharm Res* 1991; 8(suppl):S103.
96. Chowhan ZT, Chi LH. *Pharm Technol* 1985; 9:84.
97. Wand LH, Chowhan ZT. *Int J Pharm* 1990; 60:61.
98. Byrn S, et al. *Pharm Res* 1995; 9:84.
99. Byrn S, et al. *Gold Sheet* 1996; 30(6):1.
100. Byrn S, et al. Fourth International Conference on Harmonization, Brussels, July 16, 1997.
101. Drucker P. *Post-Capitalist Society.* New York: Harper Business, 1993:182.

Solutions, Emulsions, Suspensions, and Extracts

Michael M Crowley, PhD

The dosage forms described in this chapter are prepared by employing pharmaceutically and therapeutically acceptable vehicles. The active ingredient(s) may be dissolved in aqueous media, organic solvent or combination of the two, by suspending the drug (if it is insoluble) in an appropriate medium, or by incorporating the medicinal agent into one of the phases of an oil and water emulsion. Such solutions, suspensions and emulsions are further defined in subsequent paragraphs but some, with similar properties and applications, are considered in greater detail elsewhere in *Remington*.

These dosage forms are useful for a number of reasons. They can be formulated for different routes of administration: orally, introduction into body cavities, or external application. The dose can easily be adjusted by dilution, making the oral liquid form ready to be administered to children or people unable to swallow tablets or capsules. Extracts eliminate the need to isolate the drug in pure form, allow several ingredients to be administered from a single source (eg, pancreatic extract), and permit the preliminary study of drugs from natural sources. Occasionally, solutions of drugs such as potassium chloride are used to minimize adverse effects in the gastrointestinal tract.

The preparation of these dosage forms involves several considerations on the part of the pharmacist, namely; purpose of the drug, internal or external use, solubility and concentration of the drug, selection of the liquid vehicle(s), physical and chemical stability of the drug and any excipients, preservation of the preparation, and use of appropriate excipients such as buffers, solubility enhancers, suspending agents, emulsifying agents, viscosity controlling agents, colors and flavors. Oral preparations require consideration be given to improving patient compliance by making an acceptable product; consequently, color, odor and taste must be considered. The viscosity of a product also must be considered so that it has the proper palatability for an oral preparation and has the appropriate suspending properties if it is an emulsion or suspension. The theory of solutions, which involves solubility, ionization, pH control through the use of buffers, and solubilization, is discussed in Chapters 16 (*Solutions and Phase Equilibria*) and 17 (*Ionic Solutions and Electrolyte Equilibria*). Because of the complexity of some manufactured products, compounding may be carried out with the aid of linear programming models to obtain the optimal

product. Chapters 41 to 43 should be consulted for information on the preparation and characteristics of those liquid preparations that are intended for parenteral and ophthalmic use.

Much has been written about the biopharmaceutical properties of solid dosage forms. Many researchers begin their absorption studies of drugs administered in solution to assess the bioavailability relative to tablets and capsules. Absorption occurs when drugs are in a dissolved state, thus it is frequently observed that the bioavailability of oral dosage forms decreases in the following order: aqueous solution > aqueous suspension > tablet or capsule. Formulation may influence the bioavailability and pharmacokinetics of drugs in solution, including drug concentration, volume of liquid administered, pH, ionic strength, buffer capacity, surface tension, specific gravity, viscosity and excipients. Emulsions and suspensions are more complex systems; consequently, the bioavailability and pharmacokinetics of these systems may be affected by additional formulation factors such as surfactants, type of viscosity agent, particle size and particle-size distribution, polymorphism and solubility of drug in the oil phase.

Liquid preparations may be dispensed in one of three ways: (1) in its original container, (2) repackaging a bulk product at the time a prescription is presented by the patient or (3) compounding the solution, suspension, or emulsion in the dispensary. Compounding may involve nothing more than mixing marketed products in the manner indicated on the prescription or, in specific instances, may require the incorporation of active ingredients and excipients in a logical and pharmaceutically acceptable manner into aqueous or organic solvents that will form the bulk of the product.

The pharmacist, in the first instance, depends on the pharmaceutical manufacturer to produce a product that is safe, efficacious, elegant and stable until its expiration date when stored at conditions described on its label. Manufacturers guarantee efficacy of their products but, in some instances, consumer preference is variable. For example, cough syrups marketed by two different manufacturers may contain the same active ingredient(s), and the relative merits of the two products may appear interchangeable. In such instances the commercial advantage may be based on factors such as flavor, color, aroma, mouth feel and packaging.

SOLVENTS FOR LIQUID PHARMACEUTICAL PREPARATIONS

The pharmacist's knowledge of the physical and chemical characteristics of a given drug dictates the selection of the appropriate solvent for a particular formulation. In addition to solubility, solvent selection is also based on clarity, toxicity,

viscosity, compatibility with excipients, chemical inertness, palatability, odor, color, and economy. In most cases, especially solutions for oral, ophthalmic or parenteral administration, water is the preferred solvent because it meets the majority of the

above criteria better than other available solvents. Often, an auxiliary solvent is also employed to augment the solvent action of water or to contribute to a product's chemical or physical stability. Alcohol, glycerin, and propylene glycol have been frequently used for these purposes.

Solvents such as acetone, ethyl oxide, and isopropyl alcohol are too toxic for use in oral pharmaceutical preparations, but they are useful as solvents in organic chemistry and in the preparatory stages of drug development. For purposes such as this, certain solvents are officially recognized in the compendia. A number of fixed oils such as corn oil, cottonseed oil, peanut oil, and sesame oil serve useful solvent functions particularly in the preparation of oleaginous injections and are recognized in the compendia for this purpose.

WATER

The major ingredient in most of the dosage forms described herein is water. It is used both as a vehicle and as a solvent for the desired flavoring or medicinal ingredients. Its tastelessness, freedom from irritating qualities, and lack of pharmacological activity make it ideal for such purposes. There is, however, a tendency to assume that its purity is constant and that it can be stored, handled, and used with a minimum of care. Although it is true that municipal supplies must comply with Environmental Protection Agency (EPA) regulations (or comparable regulations in other countries), drinking water must be purified before it can be used in pharmaceuticals. Water quality can have a significant impact on the stability of pharmaceutical dosage forms.[1] In manufacturing environments, the design of purified water systems must meet standards outlined in the United States Pharmacopeia (USP) and be validated.[2–5]

Five of the eight solvent waters described in the USP are used in the preparation of parenterals, irrigations, or inhalations. *Purified Water* must be used for all other pharmaceutical operations, dosage forms, and, as needed, in all USP tests and assays. It must meet rigid specifications for chemical purity. Purified Water is water obtained by deionization, distillation, ion-exchange, reverse osmosis, filtration, or other suitable procedures. For parenteral administration, Water for Injection, Bacteriostatic Water for Injection, or Sterile Water for Injection must be used. Sterile water may be sterile at the time of production but may lose this characteristic if it is stored improperly.

The major impurities in water are calcium, iron, magnesium, manganese, silica, and sodium. The cations usually are combined with the bicarbonate, sulfate, or chloride anions. Hard waters are those that contain calcium and magnesium cations. Bicarbonates are the major impurity in alkaline waters. Deionization processes do not necessarily produce Purified Water that will comply with EPA requirements for drinking water. Resin columns retain phosphates and organic debris. Either alone or in combination, these substances can act as growth media for microorganisms. Observations have shown that deionized water containing 90 organisms / mL contained 10^6 organisms / mL after 24-hour storage. Ultraviolet radiant energy (240–280 nm), heat or filtration can be used to limit the growth of, kill, or remove microorganisms in water. The latter method employs membrane filters and can be used to remove bacteria from heat-labile materials.

The phenomenon of *osmosis* involves the passage of water from a dilute solution across a semi-permeable membrane to a more concentrated solution. Flow of water can be stopped by applying pressure to the concentrated solution equal to the osmotic pressure. The flow of water can be reversed by applying a pressure greater than the osmotic pressure. The process of reverse osmosis uses the latter principle; by applying pressure greater than the osmotic pressure to the concentrated solution (eg, tap water), pure water may be obtained. Organic molecules are rejected on the basis of a sieve mechanism related to their size and shape. Small organic molecules, with a molecular weight smaller than approximately 200, will pass through the membrane material. Because there are few organic molecules with a molecular weight of less than 200 in the municipal water supply, reverse osmosis usually is sufficient for the removal of organic material. The pore sizes of the selectively permeable reverse-osmosis membranes are between 0.5 and 10 nm. Viruses and bacteria larger than 10 nm are rejected if no imperfections exist in the membrane. The membranes may and do develop openings that permit the passage of microorganisms. Because of the semi-static conditions, bacteria can grow both upstream and downstream of the membrane.

ALCOHOLS

Next to water, alcohol is the second most commonly used solvent in pharmacy for many organic compounds. When mixed with water, a hydroalcoholic mixture is formed capable of dissolving both alcohol-soluble and water-soluble substances, a feature especially useful for extraction and purification of active constituents from crude drugs and synthetic procedures. Alcohol, USP, is 94.9% to 96.0% by volume, at 15.56°C of C_2H_5OH and Dehydrated Alcohol, USP, contains not less than 99.5% C_2H_5OH by volume. Dehydrated alcohol is utilized when an essentially water-free alcohol is necessary. Alcohol is widely used for its miscibility with water and its ability to dissolve many water-insoluble ingredients including drug substances, flavors, and antimicrobial preservatives. Alcohol is used in liquid products as an antimicrobial preservative or in conjunction with parabens, benzoates, sorbates, and other agents. Diluted Alcohol, NF, is prepared by mixing equal volumes of Alcohol, USP, and Purified Water, USP. Due to contraction upon mixing, the final volume of such mixtures is not the sum of the individual volumes of the two components, but is generally about 3% less.

The United States Food and Drug Administration (FDA) has expressed concern about undesired pharmacologic and potential toxic effects of alcohol when ingested by children. For this reason, manufacturers of over-the-counter (OTC) oral drug products have been asked to restrict, if possible, the use of alcohol and include appropriate warnings in the labeling. For OTC oral products intended for children under 6 years of age, the recommended alcohol content limit is 0.5%; for products intended for children 6 to 12 years of age, the recommended limit is 5%; and for products recommended for children over 12 years of age and for adults, the recommended limit is 10%.

Rubbing Alcohol, USP must be manufactured in accordance with the requirements of the US Treasury Department, Bureau of Alcohol, Tobacco, and Firearms, Formula 23-H (8 parts by volume of acetone, 1.5 parts by volume of methyl isobutyl ketone, and 100 parts by volume of ethyl alcohol). It contains not less than 68.5% and not more than 71.5% by volume of dehydrated alcohol, the remainder consisting of water and the denaturants with or without color additives and perfume oils. Rubbing Alcohol contains in each 100 mL not less than 355 mg of sucrose octaacetate or not less than 1.40 mg of denatonium benzoate. The preparation may be colored with one or more color additives listed by the FDA for use in drugs and a suitable stabilizer may be added. The use of this denaturant mixture makes the separation of ethyl alcohol from the denaturants a virtually impossible task with ordinary distillation apparatus. This discourages the illegal removal and use of the alcoholic content of rubbing alcohol as a beverage. The product is volatile and extremely flammable and should be stored in tight containers remote from ignition sources. It is used externally as a soothing rub for bedridden patients, a germicide for instruments, and a skin cleanser prior to injection.

Isopropyl Rubbing Alcohol is about 70% by volume isopropyl alcohol, the remainder consisting of water with or without color additives, stabilizers, and perfume oils. It is used

exclusively as a vehicle in topical products and applications. This preparation and a commercially available 91% isopropyl alcohol solution are commonly employed to disinfect needles and syringes for hypodermic injections of insulin and for disinfecting the skin.

Glycerin is a clear, syrupy liquid with a sweet taste and is miscible with water and alcohol. Glycerin is used in a wide variety of pharmaceutical formulations including oral, otic, ophthalmic, topical, and parenteral preparations. In topical pharmaceutical formulations and cosmetics, glycerin is used primarily for its humectant and emollient properties. In parenteral formulations, glycerin is used mainly as a solvent. In oral solutions, glycerin is used as a solvent, sweetening agent, antimicrobial preservative, and viscosity-increasing agent.

Propylene glycol has become widely used as a solvent, extractant, and preservative in a variety of liquid pharmaceutical formulations, including parenterals. Propylene glycol is a viscous liquid and is miscible with water and alcohol. It is a useful solvent with a wide range of applications and is often used in place of glycerin. As an antiseptic it is similar to ethanol, and against molds it is similar to glycerin and only slightly less effective than ethanol. Propylene glycol is also used as a carrier for emulsifiers and as a vehicle for flavors, as opposed to ethanol, due to its lack of volatility.

STABILITY CONSIDERATIONS

The stability of the active ingredient in the final product is a primary concern to the formulator. In general, drug substances are less stable in aqueous media than solid dosage forms, and it is important to properly stabilize and preserve solutions, suspensions, and emulsions that contain water. Acid–base reactions, acid or base catalysis, oxidation, and reduction can occur in these products. These reactions can arise from ingredient–ingredient interactions or container–product interactions. For pH sensitive compounds, any of these interactions may alter the pH and cause precipitation.

Vitamins, essential oils, and almost all fats and oils can be oxidized. Formulators usually use the word *auto-oxidation* when the ingredient(s) reacts with oxygen but without drastic external interference. Such reactions can be initiated by heat, light (including ultraviolet radiant energy), peroxides, or other labile compounds or heavy metals such as copper or iron. This initiation step results in the formation of a free radical that then reacts with oxygen. The free radical is regenerated and reacts with more oxygen (propagation). The reactions are terminated when the free radicals react with one another.

The effect of trace metals can be minimized by using chelating agents such as citric acid or EDTA. Antioxidants may retard or delay oxidation by rapidly reacting with free radicals as they are formed (quenching). Common antioxidants include propyl, octyl, and dodecyl esters of gallic acid, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), and the tocopherols or vitamin E. Connors and coworkers provide a detailed approach for the prevention of oxidative degradation of pharmaceuticals.[6] Common antioxidants and chelating agents used in pharmaceutical preparations are listed in Table 39-1.

The USP states that if a product must be repackaged, the container specified by the compendium must be used. For example, a suitable opaque plastic container should be used if a light-resistant container is specified. If a product is diluted, or where two products are mixed, the pharmacist should use his or her knowledge to guard against incompatibility and instability. Oral antibiotic preparations constituted into liquid form should never be mixed with other products. If the chemical stability of extemporaneously prepared liquid preparations is unknown, their use should be minimized and every care taken to ensure that product characteristics will not change during the time it must be used by the patient.

Because of the number of excipients and additives in these preparations, it is recommended all the ingredients be listed on the container to reduce the risks that confront hypersensitive patients when these products are administered. Finally, the pharmacist should inform the patient regarding the appropriate use of the product, the proper storage conditions, and the time after which it should be discarded.

PRESERVATIVES

In addition to stabilization of pharmaceutical preparations against chemical and physical degradation, liquid and semisolid preparations must be protected against microbial contamination. Nearly all products described in this chapter contain water and thus, with certain exceptions such as aqueous acids, will support microbial growth. Aqueous solutions, syrups, emulsions, and suspensions often provide excellent growth media for microorganisms such as molds, yeast, and bacteria (typically *Pseudomonas, E. coli, Salmonella,* and *Staphylococcus*).

Kurup and Wan describe many preparations that are not preserved adequately and are not able to resist microbial contamination.[7] Products such as ophthalmic and injectable preparations are sterilized by autoclaving (20 minutes at 15 pounds of pressure at 120°C followed by dry heat at 180°C for 1 hour) or filtration. However, many of them require the presence of an antimicrobial preservative to maintain aseptic conditions throughout their stated shelf life.[8] Certain hydroalcoholic and alcoholic preparations do not require addition of a chemical preservative if the alcohol content is sufficient to prevent microbial growth. In general, an alcohol content of 15% by weight in acid solutions and 18% by weight in alkaline solutions is sufficient to prevent microbial growth. Most alcohol containing preparations such as elixirs, spirits, and tinctures are self-preserving and will not require preservation. Indeed, the formulator should challenge any new preparation by procedures described in the General Tests and Assays, parts ⟨51⟩ and ⟨61⟩ of the USP and other methods reported in the literature.[9–12]

Table 39-1. Common Antioxidants and Chelating Agents Used in Liquid Pharmaceutical Dosage Forms

Antioxidants	Alpha tocopherol
	Ascorbic acid
	Acorbyl palmitate
	Butylated hydroxyanisole
	Butylated hydroxytoluene
	Monothioglycerol
	Potassium metabisulfite
	Propionic acid
	Propyl gallate
	Sodium ascorbate
	Sodium bisulfite
	Sodium metabisulfite
	Sodium sulfite
Chelating Agents	Citric acid monohydrate
	Disodium edetate
	Dipotassium edetate
	Edetic acid
	Fumaric acid
	Malic acid
	Phosphoric acid
	Sodium edetate
	Tartaric acid
	Trisodium edetate

When a preservative is required, its selection is based upon several considerations, in particular the site of use whether internal, external, or ophthalmic.[13] Several researchers have described various interactions that must be considered when preservatives are selected.[14,15] The major criteria that should be considered in selecting a preservative are as follows: It should be effective against a wide spectrum of microorganisms, stable for its shelf life, nontoxic, nonsensitizing, compatible with the ingredients in the dosage form, inexpensive, and relatively free of taste and odor.

The chosen preservative should be sufficiently stable and soluble to achieve adequate concentration to provide protection. This choice is more critical in two and three phase emulsion systems in which the preservative may be more soluble in the oil phase than in the aqueous phase.[12,16] The pH of the preparation must be considered to ensure that the preservative does not dissociate rendering it ineffective or degrade by acid or base catalyzed hydrolysis. The undissociated moiety or molecular form of a preservative possesses preservative capacity because the ionized form is unable to penetrate microorganisms. The preservative must be compatible with the formulation ingredients and the product container or closure. Finally, the preservative must not impact the safety or comfort of the patient when administered. For instance, preservatives used in ophthalmic preparations must be non-irritating. Chlorobutanol, benzalkonium chloride, and phenylmercuric nitrate are commonly used in these applications.

Although few microorganisms are viable below a pH of 3 or above pH 9, most aqueous pharmaceutical preparations are manufactured within the favorable pH range. Acidic preservatives such as benzoic acid, boric acid, and sorbic acid are less dissociated and more effective in acidic formulations. Similarly, alkaline preservatives are less effective in acidic or neutral conditions and more effective in alkaline formulations. The scientific literature is rife with examples of incompatibilities between preservatives and other pharmaceutical adjuncts.[17–19] Commonly used macromolecules including cellulose derivatives, polyethylene glycol and tragacanth gum have been reported to cause preservative failure due to binding and adsorption.[20,21]

The mode of action by which preservatives interfere with microbial growth, multiplication, and metabolism occurs through one of several mechanisms. Preservatives often alter cell membrane permeability causing leakage of cell constituents (partial lysis), complete lysis, and cytoplasmic leakage and / or coagula-

tion of cytoplasmic constituents (protein precipitation). Other preservatives inhibit cellular metabolism by interference with enzyme systems or cell wall synthesis, oxidation of cellular constituents, or hydrolysis.

Preservatives commonly used in pharmaceutical products are listed in Table 39-2 with typical concentration levels. Preservatives may be grouped into a number of classes depending upon their molecular structure. These basic groups are discussed below.

Alcohols

Ethanol is useful as a preservative when it is used as a solvent; however, it does need a relatively high concentration, somewhat greater than 15%, to be effective. Too high a concentration may result in incompatibilities in suspension and emulsion systems. Propylene glycol also is used as a solvent in oral solutions and topical preparations, and it can function as a preservative in the range of 15% to 30%. It is not volatile like ethanol and is used frequently not only in solutions but also in suspensions and emulsions. Chlorobutanol and phenylethyl alcohol are other alcohols used in lower concentrations (about 1%) as preservatives.

Acids

Benzoic acid has a low solubility in water, about 0.34% at 25°C, but the apparent aqueous solubility of benzoic acid may be enhanced by the addition of citric acid or sodium acetate to the solution. The concentration range used for inhibitory action varies from 0.1% to 0.5%. Activity depends on the pH of the medium because only the undissociated acid has antimicrobial properties. Optimum activity occurs at pH values below 4.5; at values above pH 5, benzoic acid is almost inactive.[22] It has been reported that antimicrobial activity of benzoic acid is enhanced by the addition of the basic protein protamine.[23] Sorbic acid also has a low solubility in water, 0.3% at 30°C. Suitable concentrations for preservative action are in the range of 0.05 to 2%. Its preservative action is due to the nonionized form; consequently, it is only effective in acid media. The optimum antibacterial activity is obtained at pH 4.5, and practically no activity is observed above pH 6. Sorbic acid is subject to oxidation, particularly in the presence of light and in aqueous

Table 39-2. Common Preservatives Used in Liquid Pharmaceutical Dosage Forms and Their Typical Concentration Levels

ANTIMICROBIAL PRESERVATIVES	TYPICAL USAGE LEVEL (% W/W)	ANTIFUNGAL PRESERVATIVES	TYPICAL USAGE LEVEL (% W/W)
Benzalkonium Chloride	0.002–0.02%	Butyl Paraben	0.1–0.4%
Benzethonium Chloride	0.01–0.02%	Methyl Paraben	0.1–0.25%
Benzyl Alcohol	3.0%	Ethyl Paraben	0.1–0.25%
Bronopol	0.01–0.1%	Propyl Paraben	0.1–0.25%
Cetrimide	0.005%	Benzoic Acid	0.1–0.5%
Cetylpyridinium chloride	0.0005–0.0007%	Potassium sorbate	0.1–0.2%
Chlorhexidine	0.002–0.5%	Sodium Benzoate	0.1–0.2%
Chlorobutanol	0.5%	Sodium Propionate	5–10%
Chlorocresol	0.2%	Sorbic Acid	0.05–0.2%
Chloroxylenol	0.1–0.8%		
Cresol	0.15–0.3%		
Ethyl Alcohol	15–20%		
Glycerin	20–30%		
Hexetidine	0.1%		
Imidurea	0.03–0.5%		
Phenol	0.1–0.5%		
Phenoxyethanol	0.5–1.0%		
Phenylethyl Alcohol	0.25–0.5%		
Phenylmercuric Nitrate	0.002–0.01%		
Propylene Glycol	15–30%		
Thimerosal	0.1%		

solutions. Activity against bacteria can be variable because of its limited stability. Thus, sorbic acid is frequently used in combination with other antimicrobial preservatives or glycols in which synergistic effects occur.

Esters

Parabens are esters of *p*-hydroxybenzoic acid and include the methyl, ethyl, propyl, and butyl derivatives. The water solubility of the parabens decreases as the molecular weight increases from 0.25% for the methyl ester to 0.02% for the butyl ester. These compounds are used widely in pharmaceutical products, stable over a pH range of 4 to 8, and have a broad spectrum of antimicrobial activity, although they are most effective against yeasts and molds. Antimicrobial activity increases as the chain length of the alkyl moiety is increased, but aqueous solubility decreases; therefore, a mixture of parabens is frequently used to provide effective preservation. Preservative efficacy is also improved by the addition of propylene glycol (2–5%) or by using parabens in combination with other antimicrobial agents such as imidurea. Activity is reduced in the presence of nonionic surface active agents due to binding. In alkaline solutions, ionization takes place and this reduces their activity; in addition, hydrolytic decomposition of the ester group occurs with a loss of activity.

Quaternary Ammonium Compounds

Benzalkonium chloride is a mixture consisting principally of the homologs $C_{12}H_{25}$ and $C_{14}H_{29}$. This preservative is used at a relatively low concentration, 0.002% to 0.02%, depending on the nature of the pharmaceutical product. This class of compounds has an optimal activity over the pH range of 4 to 10 and is quite stable at room temperature. Because of the cationic nature of this type of preservative, it is incompatible with many anionic compounds such as surfactants and can bind to nonionic surfactants. It is used generally in preparations for external use or those solutions that come in contact with mucous membranes. In ophthalmic preparations, benzalkonium chloride is widely used at a concentration of 0.01–0.02% w/w. Often it is used in combination with other preservatives or excipients, particularly 0.1% w/v disodium edetate, to enhance its antimicrobial activity against strains of *Pseudomonas*. A concentration of 0.002–0.02% is used in nasal and otic formulations, sometimes in combination with 0.002–0.005% thimerosal. Benzalkonium chloride 0.01% w/v is also employed as a preservative in small-volume parenteral products.

Clearly, when the pharmacist dispenses or compounds liquid preparations, responsibility is assumed, along with the manufacturer, for the maintenance of product stability. General chapter ⟨1191⟩ of the USP describes stability considerations for dispensing, which should be studied in detail.[9] Stock should be rotated and replaced if expiration dates on the label so indicate. Products should be stored in the manner indicated on the manufacturer's label or in the compendium. Further, products should be checked for evidence of instability. With respect to solutions, elixirs, and syrups, major signs of instability are color change, precipitation, and evidence of microbial or chemical gas formation. Emulsions may cream, but if they break (ie, there is a separation of an oil phase) the product is considered unstable. Sedimentation and caking are primary indications of instability in suspensions. The presence of large particles may mean that excessive crystal growth has occurred (Ostwald Ripening). Additional details on these topics are provided in the pertinent sections of this chapter.

SOLUTIONS

A solution is a homogeneous mixture that is prepared by dissolving a solid, liquid, or gas in another liquid and represents a group of preparations in which the molecules of the solute or dissolved substance are dispersed among those of the solvent. Most solutions are unsaturated with the solute, in other words, the concentration of the solute in the solution is below its solubility limit. The strengths of pharmaceutical solutions are usually expressed in terms of % strength, although for very dilute preparations expressions of ratio strength are sometimes used. The term % when used without qualification (as with w/v, v/v, or w/w) means % weight-in-volume for solutions or suspensions of solids in liquids; % weight-in-volume for solutions of gases in liquids; % volume-in-volume for solutions of liquids in liquids; and weight-in-weight for mixtures of solids and semisolids.

Solutions also may be classified on the basis of physical or chemical properties, method of preparation, use, physical state, number of ingredients, and particle size. For the pharmacist, solutions are more defined by site of administration and composition than by physicochemical definitions. For instance, pharmaceutical solutions may be classified as an *oral solution, otic solution, ophthalmic solution,* or *topical solution*. These solutions may also be classified based upon their composition. *Syrups* are aqueous solutions containing a sugar; *elixirs* are sweetened hydroalcoholic (combinations of water and ethanol) solutions; *spirits* are solutions of aromatic materials if the solvent is alcoholic or *aromatic waters* if the solvent is aqueous. Depending on their method of preparation and concentration, *tinctures* or *fluid extracts* are solutions prepared by extracting active constituents from crude drugs.

Many pharmaceutical chemicals are only slowly soluble in a given solvent and require an extended time for complete dissolution. To increase the dissolution rate, a pharmacist may employ one or several techniques such as applying heat, reducing the particle size of the solute, utilizing of a solubilizing agent, or subjecting the ingredients to rigorous agitation. In most cases, solutes are more soluble in solvents at elevated temperatures than at room temperature or below due to the endothermic nature of the dissolution process. The pharmacist should ensure that the materials are heat stable and non-volatile when using heat to facilitate the dissolution rate.

AQUEOUS SOLUTIONS

The narrower definition in this subsection limits the solvent to water and excludes those preparations that are sweet and/or viscid in character and nonaqueous solutions. This section includes those pharmaceutical forms that are designated as *Aromatic Waters, Aqueous Acids, Solutions, Douches, Enemas, Gargles, Mouthwashes, Juices, Nasal Solutions, Otic Solutions,* and *Irrigation Solutions*.

Aromatic Waters

The USP defines Aromatic Waters as clear, saturated aqueous solutions (unless otherwise specified) of volatile oils or other aromatic or volatile substances.[9] Their odors and tastes are similar, respectively, to those of the drugs or volatile substances from which they are prepared, and they are free from empyreumatic and other foreign odors. Aromatic waters may be prepared by distillation or solution of the aromatic substance, with or without the use of a dispersing agent. They are used principally as flavored or perfumed vehicles.

Peppermint Water USP and Stronger Rose Water USP are examples of aromatic waters. Concentrated waters, such as peppermint, dill, cinnamon, and caraway, may be prepared as follows:

Dissolve 20 mL of the volatile oil in 600 mL of 90% ethanol. Add sufficient purified water in successive small portions to produce 1000 mL. Shake vigorously after each addition. Add 50 g of sterilized purified talc, shake occasionally for several hours, and filter.

The aromatic water is prepared by diluting the concentrate with 39 times its volume of water.

The chemical composition of many of the volatile oils is known, and suitable synthetic substances may be used in preparing pharmaceuticals and cosmetics. Similarly, many synthetic aromatic substances have a characteristic odor; for example, geranyl phenyl acetate has a honey odor. Such substances, either alone or in combination, can be used in nonofficial preparations.

The principal difficulty experienced in compounding prescriptions containing aromatic waters is *salting out* certain ingredients such as very soluble salts. A replacement of part of the aromatic water with purified water is permissible when no other function is being served than that of a vehicle. Aromatic waters will deteriorate with time and should, therefore, be made in small quantities, protected from intense light and excessive heat, and stored in airtight, light-resistant containers.

Aqueous Acids

Inorganic acids and certain organic acids, although of minor significance as therapeutic agents, are of great importance in pharmaceutical manufacturing and analysis. This is especially true of acetic, hydrochloric, and nitric acids. Many of the more important inorganic acids are available commercially in the form of concentrated aqueous solutions. The percentage strength varies from one acid to another and depends on the solubility and stability of the solute in water and on the manufacturing process. Thus, Hydrochloric Acid contains from 36.5% to 38.0% by weight of HCl, whereas Nitric Acid contains from 69% to 71% by weight of HNO_3.

Because the strengths of these concentrated acids are stated in terms of percent by weight, it is essential that specific gravities also be provided if one is to be able to calculate conveniently the amount of absolute acid contained in a unit volume of the solution as purchased. The mathematical relationship involved is given by the equation $M = V \times S \times F$, where M is the mass in g of absolute acid contained in V mL of solution having a specific gravity S and a fractional percentage strength F.

As an example, Hydrochloric Acid containing 36.93% by weight of HCl has a specific gravity of 1.1875. Therefore, the amount of pure HCl supplied by 100 mL of this solution is given by:

$$M = 100 \times 1.1875 \times 0.3693 = 43.85 \text{ g HCl}$$

Although many of the reactions characteristic of acids offer opportunities for incompatibilities, only a few are of sufficient importance to require more than casual mention. Acids and acid salts decompose carbonates with liberation of carbon dioxide; in a closed container, sufficient pressure may be developed to produce an explosion. Inorganic acids react with salts of organic acids to produce the free organic acid and a salt of the inorganic acid. If insoluble, the organic acid will be precipitated. Thus, salicylic acid and benzoic acid are precipitated from solutions of salicylates and benzoates. Boric acid likewise is precipitated from concentrated solutions of borates. By a similar reaction, certain soluble organic compounds are converted into an insoluble form. Phenobarbital sodium, for example, is converted into phenobarbital that will precipitate in aqueous solution.

The ability of acids to combine with alkaloids and other organic compounds containing a basic nitrogen atom is used in preparing soluble salts of these substances. Certain solutions, syrups, elixirs, and other pharmaceutical preparations, may contain free acid, which causes these preparations to exhibit the incompatibilities characteristic of the acid. Acids also possess the incompatibilities of the anions that they contain and, in the case of organic acids, these are frequently of prime importance. These are discussed under the specific anions.

Diluted Acids

The diluted acids in the USP are aqueous solutions of acids of a suitable strength (usually 10% w/v but Diluted Acetic Acid is 6% w/v) for internal administration or for the manufacture of other preparations.

The strengths of the official undiluted acids are expressed as percentages in weight (w/w), whereas the strengths of the official diluted acids are expressed as percent in volume (w/v). It, therefore, becomes necessary to consider the specific gravities of the concentrated acids when calculating the volume required to make a given quantity of diluted acid. The following equation will give the number of milliliters required to make 1000 mL of diluted acid:

$$\frac{\text{Strength of diluted acid} \times 1,000}{\substack{\text{Strength of undiluted acid} \\ \times \text{ Specific gravity of undiluted acid}}}$$

Thus, if one wishes to make 1000 mL of Diluted Hydrochloric Acid USP (10% w/v) using Hydrochloric Acid that assays 37.5% HCl (sp gr 1.18), the amount required is

$$\frac{10 \times 1,000}{37.5 \times 1.18} = 226 \text{ mL}$$

Diluted Hydrochloric Acid, USP has been used in the treatment of achlorhydria. However, it may irritate the mucous membrane of the mouth and attack the enamel of the teeth. The usual dose is 2 to 4 mL, well-diluted with water. In the treatment of achlorhydria no attempt is made to administer more than a relief-producing dose.

Douches

A douche is an aqueous solution directed against a part or into a cavity of the body. It functions as a cleansing or antiseptic agent. An *eye douche*, used to remove foreign particles and discharges from the eyes, is directed gently at an oblique angle and allowed to run from the inner to the outer corner of the eye. *Pharyngeal douches* are used to prepare the interior of the throat for an operation and cleanse it in suppurative conditions. Similarly, there are *nasal douches* and *vaginal douches*. Douches usually are directed to the appropriate body part by using bulb syringes.

Douches are often dispensed in the form of a powder with directions for dissolving in a specified quantity of water (usually warm). However, tablets for preparing solutions are available (eg, Dobell's Solution Tablets) or the solution may be prepared by the pharmacist. If powders or tablets are supplied, they must be free from insoluble material in order to produce a clear solution. Tablets are produced by the usual processes but any lubricants or diluents used must be readily soluble in water. Boric acid may be used as a lubricant and sodium chloride normally is used as a diluent. Tablets deteriorate on exposure to moist air and should be stored in airtight containers.

Douches are not official as a class of preparations but several substances in the compendia frequently are employed as such in weak solutions. *Vaginal douches* are the most common type of douche and are used for cleansing the vagina and hygienic purposes. Liquid concentrates or powders, which may be prepared in bulk or as single-use packages, should be diluted or dissolved in the appropriate amount of warm water prior to use.

The ingredients used in vaginal douches include antimicrobial agents such as benzalkonium chloride, the parabens or chlorothymol, and anesthetics or antipruritics such as phenol or menthol. Astringents such as zinc sulfate or potassium alum, surface-active agents such as sodium lauryl sulfate, and chemicals to alter the pH such as sodium bicarbonate or citric acid also are used.

Enemas

A number of solutions are administered rectally for the local effects of the medication (eg, hydrocortisone) or for systemic absorption (eg, aminophylline). In the case of aminophylline, the rectal route of administration minimizes the undesirable gastrointestinal reactions associated with oral therapy.[24] Clinically effective blood levels of the agents are usually obtained within 30 minutes following rectal instillation. Corticosteroids are administered as retention enemas or continuous drip as adjunctive treatment of some patients with ulcerative colitis.

Enema preparations are rectal injections employed to evacuate the bowel (evacuation enemas), influence the general system by absorption, or to affect a local disease. The latter two are called retention enemas. They may possess anthelmintic, nutritive, sedative, or stimulating properties, or they may contain radiopaque substances for roentgenographic examination of the lower bowel.

Sodium chloride, sodium bicarbonate, sodium monohydrogen phosphate, sodium dihydrogen phosphate, glycerin, docusate potassium, and light mineral oil are used in enemas to evacuate the bowel. These substances may be used alone, in combination with each other, or in combination with irritants such as soap. Evacuation enemas usually are given at body temperature in quantities of 1 to 2 pt injected slowly with a syringe.

An official retention enema used for systemic purposes is aminophylline. Retention enemas are to be retained in the intestine and should not be used in larger quantities than 150 mL for an adult. Usually, the volume is considerably smaller, such as a few mL. *Microenema* is a term used to describe these small-volume preparations. Vehicles for retention microenemas have been formulated with small quantities of ethanol and propylene glycol, and no significant difference in irritation, as compared with water, was found. A number of other drugs such as valproic acid, indomethacin, and metronidazole have been formulated as microenemas for the purpose of absorption.

Gargles

Gargles are aqueous solutions frequently containing antiseptics, antibiotics, and/or anesthetics used for treating the pharynx and nasopharynx by forcing air from the lungs through the gargle that is held in the throat; subsequently, the gargle is expectorated. Many gargles must be diluted with water prior to use. Although mouthwashes are considered as a separate class of pharmaceuticals, many are used as gargles either as is, or diluted with water.

A gargle/mouthwash containing the antibiotic tyrothricin has been shown to provide levels of gramicidin, a component of tyrothricin, in saliva when used as a gargle rather than a mouthwash.[25] Higher saliva levels of gramicidin were obtained when a lozenge formulation was employed. Rapid relief of pharyngeal and oral pain was obtained when Cepacaine solution, which contains a topical anesthetic, was used as a gargle.[26]

Nystatin is administered in both powder and liquid form to treat oral fungal infections.[27] The medication is taken by placing one-half of the dose in each side of the mouth, swishing it around as long as possible, then gargling and swallowing. Hydrogen peroxide is a source of nascent oxygen and a weak topical antibacterial agent. Hydrogen peroxide topical solution has been used as a mouthwash or gargle in the treatment of pharyngitis or Vincent's stomatitis.[28,29] Hydrogen peroxide has

also been applied in root canals of teeth or other dental pulp cavities. While used topically as a 1.5–3% solution for cleansing wounds, hydrogen peroxide is usually diluted with an equal volume of water for use as a mouthwash or gargle. Hydrogen peroxide gel is used topically as a 1.5% gel for cleansing minor wounds or irritations of the mouth or gums. A small amount of the gel is applied to the affected area, allowed to remain in place for at least 1 minute, and then expectorated; the gel may be used up to 4 times daily (after meals and at bedtime).

Mouthwashes

Mouthwashes are aqueous solutions often in concentrated form containing one or more active ingredients and excipients described below. They are used by swishing the liquid in the oral cavity. Mouthwashes can be used for two purposes, therapeutic and cosmetic. Therapeutic rinses or washes can be formulated to reduce plaque, gingivitis, dental caries, and stomatitis. Cosmetic mouthwashes may be formulated to reduce bad breath through the use of antimicrobial and/or flavoring agents.

Recent information indicates that mouthwashes are being used as a dosage form for a number of specific problems in the oral cavity; for example, mouthwashes containing a combination of antihistamines, hydrocortisone, nystatin, and tetracycline have been prepared from commercially available suspensions, powders, syrups, or solutions for the treatment of stomatitis, a painful side effect of cancer chemotherapy. Other drugs include allopurinol, also used for the treatment of stomatitis,[30] pilocarpine for xerostoma (dry mouth),[31] amphotericin B for oral candidiasis,[32] and chlorhexidine gluconate for plaque control.[33] Mouthwashes may be used for diagnostic purposes. For example, oral cancer and lesions are detected using toluidine blue mouth rinse.[34]

Commercial products (eg, Cepacol, Listerine, Micrin, or Scope) vary widely in composition. Tricca has described the excipients generally found in Mouthwashes as alcohols, surfactants, flavors, and coloring agents.[35] Alcohol is often present in the range of 10% to 20%. It enhances the flavor, provides sharpness to the taste, aids in masking the unpleasant taste of active ingredients, functions as a solubilizing agent for some flavoring agents, and may function as a preservative. Humectants such as glycerin and sorbitol may form 5% to 20% of the mouthwash. These agents increase the viscosity of the preparation and provide a certain *body* or *mouth feel* to the product. They enhance the sweetness of the product and, along with the ethanol, improve the preservative qualities of the product.

Surfactants of the nonionic class such as polyoxyethylene/polyoxypropylene block copolymers or polyoxyethylene derivatives of sorbitol fatty acid esters may be used. The concentration range is 0.1% to 0.5%. An anionic surfactant occasionally used is sodium lauryl sulfate. Surfactants are used because they aid in the solubilization of flavors and in the removal of debris by providing foaming action. Cationic surfactants such as cetylpyridinium chloride are used for their antimicrobial properties, but these tend to impart a bitter taste.

Flavors are used in conjunction with alcohol and humectants to overcome disagreeable tastes, at the same time flavors must be safe to use. The principle flavoring agents are peppermint, spearmint, cinnamon, wintergreen oils, menthol, or methyl salicylate. Other flavoring agents may be used singly or in combination. Finally, coloring agents also are used in these products.

Juices

A juice is prepared from fresh ripe fruit, is aqueous in character, and is used in making syrups that are employed as vehicles. The freshly expressed juice is preserved with benzoic acid and allowed to stand at room temperature for several days, until the pectins that naturally are present are destroyed by enzymatic action, as indicated by the filtered juice yielding a clear solution

with alcohol. Pectins, if allowed to remain, would cause precipitation in the final syrup.

Cherry Juice and Tomato Juice are described in the USP. Artificial flavors now have replaced many of the natural fruit juices. Although they lack the flavor of the natural juice, they are more stable and easier to incorporate into the final pharmaceutical form. Commercial juices such as orange, apple, grape, and mixed vegetables have been used recently to prepare extemporaneous preparations of cholestyramine[36] and nizatidine.[37] Information on cranberry juice indicates that it may be effective in controlling some urinary tract infections and urolithiasis.[38]

Nasal Solutions

Nasal solutions are usually aqueous solutions designed to be administered to the nasal passages in drops or sprays. Other nasal preparations may be in the form of emulsions or suspensions. The adult nasal cavity has about a 20 mL capacity with a large surface area (about 180 cm^2) for drug absorption afforded by the microvilli present along the pseudo-stratified columnar epithelial cells of the nasal mucosa.[39] The nasal tissue is highly vascularized making it an attractive site for rapid and efficient systemic absorption. Another advantage of nasal delivery is that it avoids first-pass metabolism by the liver. For some peptides and small molecular compounds, intranasal bioavailability has been comparable to that of injections. However, bioavailability decreases as the molecular weight of a compound increases, and for proteins composed of more than 27 amino acids bioavailability may be low.[40] Various pharmaceutical techniques and functional excipients, such as surfactants, have been shown to be capable of enhancing the nasal absorption of large molecules.[41,42]

Many drugs are administered for their local sympathomimetic effects to reduce nasal congestion, such as Ephedrine Sulfate Nasal Solution, USP or Naphazoline Hydrochloride Nasal Solution, USP. A few other preparations, Lypressin Nasal Solution USP and Oxytocin Nasal Solution USP, are administered in spray form for their systemic effect for the treatment of diabetes insipidus and milk letdown prior to breast feeding, respectively. Examples of commercial products for nasal use are listed in Table 39-3.

Nasal solutions are formulated to be similar to nasal secretions with regard to toxicity, pH, and viscosity so that normal ciliary action is maintained. Thus, aqueous nasal solutions usually are isotonic and slightly buffered to maintain a pH of 5.5 to 6.5. In addition, antimicrobial preservatives, similar to those used in ophthalmic preparations, and appropriate drug stabilizers, if required, are included in the formulation.

Current studies indicate that nasal sprays are deposited mainly in the atrium and cleared slowly into the pharynx with the patient in an upright position. Drops spread more extensively than the spray, and three drops cover most of the walls of the nasal cavity with the patient in a supine position and head tilted back and turned left and right.[43,44] It is suggested that drop delivery, with appropriate movement by the patient, leads to extensive coverage of the walls of the nasal cavity.

Most nasal solutions are packaged in dropper or spray bottles, usually containing 15 to 30 mL of medication. The formulator should ensure the product is stable in the containers and the pharmacist should keep the packages tightly closed during periods of nonuse. The patient should be advised that should the solution become discolored or contain precipitated matter, it must be discarded.

Otic Solutions

These solutions occasionally are referred to as ear or aural preparations. Other otic preparations include suspensions and ointments for topical application in the ear. Ear preparations are usually placed in the ear canal by drops or in small amounts for the removal of excessive cerumen (ear wax) or for the treatment of ear infections, inflammation, or pain.

The main classes of drugs used for topical administration to the ear include analgesics, such as benzocaine; antibiotics, such as neomycin; and anti-inflammatory agents, such as cortisone (Table 39-4). The USP preparations include Antipyrine and Benzocaine Otic Solution. The Neomycin and Polymyxin B Sulfates and Hydrocortisone Otic Solutions may contain appropriate buffers, solvents, and dispersants usually in an aqueous solution. The main solvents used in these preparations include glycerin or water. The viscous glycerin vehicle permits the drug to remain in the ear for a long time. Anhydrous glycerin, being hygroscopic, tends to remove moisture from surrounding tissues, thus reducing swelling. Viscous liquids such as glycerin or propylene glycol are used either alone or in combination with a surfactant to aid in the removal of cerumen (ear wax). To provide sufficient time for aqueous preparations to act, it is necessary for patients to remain on their side for a few minutes so the drops do not run out of the ear. Otic preparations are dispensed in a container that permits the administration of drops.

Irrigation Solutions

Irrigation solutions are sterile, non-pyrogenic solutions used to wash or bathe surgical incisions, wounds, or body tissues. Because they come in contact with exposed tissue, they must meet stringent USP requirements for sterility, total solids, and bacterial endotoxins. These products may be prepared by dissolving the active ingredient in Water for Injection. They are packaged in single-dose containers, preferably Type I or Type II glass, or suitable plastic containers, and then sterilized. A number of irrigations are described in the USP, including Acetic

Table 39-3. Examples of Commercial Nasal Preparations

PRODUCT NAME	MANUFACTURER	ACTIVE INGREDIENT	INDICATION
Atrovent Nasal Spray	Boehringer Ingelheim	Ipratropium bromide 0.06%	Seasonal or Allergic Rhinitis
Beconase AQ Nasal Spray	GlaxoSmithKline	Beclomethasone dipropionate, monohydrate 42 mcg	Seasonal or Allergic Rhinitis
Miacalcin	Novartis	Calcitonin-salmon, 2200 I.U. per mL	Postmenopausal osteoporosis
Nasalcrom Nasal Spray	Pharmacia	Cromolyn sodium 5.2 mg	Seasonal or Allergic Rhinitis
Nasarel Nasal Spray	IVAX	Flunisolide	Seasonal or perennial rhinitis
Nicotrol Nasal Spray	Pfizer	Nicotine 0.5 mg	Smoking Cessation
Neo-Synephrine	Bayer	Oxymetazoline hydrochloride 0.05%	Decongestion
Rhinocort Aqua Nasal Spray	Astra-Zeneca	Budesonide 32mcg	Seasonal or Allergic Rhinitis
Stadol Nasal Spray	Bristol-Myers Squibb	Butorphanol tartrate, 1 mg	Pain Relief, Migraines
Stimate Nasal Spray	Aventis	Desmopressin Acetate 1.5 mg/mL	Hemophilia A or von Willebrand disease
Synare Nasal Solution	Searle	Nafarelin acetate 2 mg/mL	Endometriosis
Tyzine	Bradley Pharmaceuticals	Tetrahydrozoline hydrochloride	Decongestion

Table 39-4. Examples of Commercial Otic Preparations

PRODUCT NAME	MANUFACTURER	ACTIVE INGREDIENT	INDICATION
Americaine-Otic	Celltech	Benzocaine	Local anesthetics
Cerumenex Ear Drops	Purdue	Triethanolamine polypeptide oleate-condensate	Removal of earwax
Chloromycetin Otic	Pfizer	Chloramphenicol	Antiinfective
Cipro HC Otic	Alcon	Ciprofloxacin hydrochloride and hydrocortisone	Acute otitis externa
Cortisporin	GlaxoSmithKline	Neomycin and Polymyxin B Sulfates and Hydrocortisone	Antibacterial and anti-inflammatory
Debrox Drops	GlaxoSmithKline	Carbamide peroxide	Removal of earwax
Floxin Otic	Daiichi	Ofloxacin	Antiinfective
Tympagesic	Savage	Antipyrine, Benzocaine, and Phenylephrine Hydrochloride	Topical anesthetic

Acid Irrigation for bladder irrigation, Dimethyl Sulfoxide Irrigation for relief of internal cystitis, Glycine Irrigation for transurethral prostatic resection, Ringer's Irrigation for general irrigation, Neomycin and Polymyxin B Sulfates Solution for Irrigation for infection, and Sodium Chloride Irrigation for washing wounds.

Extemporaneous formulations frequently are prepared using an isotonic solution of sodium chloride as the solvent. For example, cefazolin or gentamicin in 0.9% sodium chloride are used as anti-infective irrigations[45] and 5-fluororacil in 0.9% sodium chloride is employed for bladder irrigation.[46] Alum, either potassium or ammonium, in either sterile water or 0.9% sodium chloride for irrigation has been used for bladder hemorrhage. Amphotericin in sterile water has been used for the treatment of localized infections on the dermis, the bladder, and urinary tract.[47] All the extemporaneous preparations should meet the general requirements noted above for USP irrigations.

PREPARATION OF SOLUTIONS

The method of preparation for many solutions is given in the compendia. These procedures fall into three main categories: simple solutions, solution by chemical reaction, and solution by extraction.

Simple Solutions are prepared by dissolving the solute in most of the solvent, mixing until dissolved, then adding sufficient solvent to bring the solution up to the proper volume. The solvent may contain other ingredients that stabilize or solubilize the active ingredient. Calcium Hydroxide Topical Solution USP (Lime Water), Sodium Phosphates Oral Solution USP, and Strong Iodine Solution USP are examples.

Calcium Hydroxide Topical Solution USP contains, in each 100 mL, not less than 140 mg of $Ca(OH)_2$. The solution is prepared by agitating vigorously 3 g of calcium hydroxide with 1000 mL of cool, purified water. Excess calcium hydroxide is allowed to settle out and the clear, supernatant liquid dispensed. An increase in solvent temperature usually implies an increase in solute solubility. This rule does not apply, however, to the solubility of calcium hydroxide in water, which decreases with increasing temperature. The official solution is prepared at 25°C.

Solutions containing hydroxides react with the carbon dioxide in the atmosphere.

$$OH^- + CO_2 \rightarrow HCO_3^-$$

$$OH^- + HCO_3^- \rightarrow CO_3{}^{2-} + H_2O$$
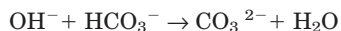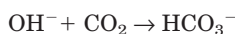
Calcium Hydroxide Topical Solution, therefore, should be preserved in well-filled, tight containers, at a temperature not exceeding 25°C.

Strong Iodine Solution USP contains, in each 100 mL, 4.5 to 5.5 g of iodine, and 9.5 to 10.5 g of potassium iodide. It is prepared by dissolving 50 g of iodine in 100 mL of purified water

containing 100 g of potassium iodide. Sufficient purified water then is added to make 1000 mL of solution. One g of iodine dissolves in 2950 mL of water. However, solutions of iodides dissolve large quantities of iodine. Strong Iodine Solution is, therefore, a solution of polyiodides in excess iodide.

$$I^- + nI_2 \rightarrow I_{(2n + 1)}{}^-$$
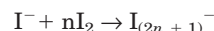
Doubly charged anions may be found also.

$$2I^- + nI_2 \rightarrow I_{(2n + 2)}{}^{2-}$$
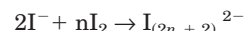
Strong Iodine Solution is used in the treatment of iodide deficiency disorders such as endemic goiter.
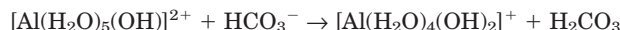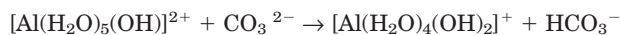
Several antibiotics (eg, cloxacillin sodium, nafcillin sodium, and vancomycin), because they are relatively unstable in aqueous solution, are prepared by manufacturers as dry powders or granules in combination with suitable buffers, colors, diluents, dispersants, flavors, and/or preservatives. These preparations, Cloxacillin Sodium for Oral Solution, Nafcillin for Oral Solution, and Vancomycin Hydrochloride for Oral Solution meet the requirements of the USP. Immediately prior to dispensing to the patient, the pharmacist adds the appropriate amount of water. The products are stable for up to 14 days when refrigerated.[48] This period usually provides sufficient time for the patient to complete the administration of all the medication.

Solutions by chemical reaction are prepared by reacting two or more solutes with each other in a suitable solvent. An example is Aluminum Subacetate Topical Solution USP. Aluminum sulfate (145 g) is dissolved in 600 mL of cold water. The solution is filtered, and precipitated calcium carbonate (70 g) is added, in several portions, with constant stirring. Acetic acid (160 mL) is added slowly and the mixture set aside for 24 hours. The product is filtered and the magma on the Buchner filter washed with cold water until the total filtrate measures 1,000 mL.

The solution contains pentaquohydroxo- and tetraquodihydroxoaluminum(III) acetates and sulfates dissolved in an aqueous medium saturated with calcium sulfate. The solution contains a small amount of acetic acid. It may be stabilized by the addition of not more than 0.9% boric acid. The reactions involved in the preparation of the solution are given below. The hexaquo aluminum cations first are converted to the nonirritating

$$[Al(H_2O)_5(OH)]^{2+} \text{ and } [Al(H_2O)_4(OH)_2]^+ \text{ cations.}$$

$$[Al(H_2O)_6]^{3+} + CO_3{}^{2-} \rightarrow [Al(H_2O)_5(OH)]^{2+} + HCO_3^-$$

$$[Al(H_2O)_6]^{3+} + HCO_3^- \rightarrow [Al(H_2O)_5(OH)]^{2+} + H_2O + CO_2$$

As the concentration of the hexaquo cations decreases, secondary reactions involving carbonate and bicarbonate occur.

$$[Al(H_2O)_5(OH)]^{2+} + CO_3{}^{2-} \rightarrow [Al(H_2O)_4(OH)_2]^+ + HCO_3^-$$

$$[Al(H_2O)_5(OH)]^{2+} + HCO_3^- \rightarrow [Al(H_2O)_4(OH)_2]^+ + H_2CO_3$$

The pH of the solution now favors the precipitation of dissolved calcium ions as the insoluble sulfate. Acetic acid now is added.

The bicarbonate that is formed in the final stages of the procedure is removed as carbon dioxide.

Aluminum Subacetate Topical Solution is used in the preparation of Aluminum Acetate Topical Solution USP (Burow's Solution). The latter solution contains 15 mL of glacial acetic acid, 545 mL of Aluminum Subacetate Topical Solution and sufficient water to make 1000 mL. It is defined as a solution of aluminum acetate in approximately 5%, by weight, of acetic acid in water. It may be stabilized by the addition of not more than 0.6% boric acid.

Often, drugs or pharmaceutical necessities of vegetable or animal origin often are extracted with water or with water containing other substances. Preparations of this type may be classified as solutions but, more often, are classified as extracts and are described at the end of this chapter.

SWEET AND OTHER VISCID AQUEOUS SOLUTIONS

Solutions that are sweet or viscid include syrups, honeys, mucilages, and jellies. All of these are viscous liquids or semisolids. The basic sweet or viscid substances giving body to these preparations are sugars, polyols, and / or polysaccharides.

Syrups

Syrups are concentrated, viscous, aqueous solutions of sugar or a sugar substitute with or without flavors and medical substances. When Purified Water alone is used in making the solution of sucrose, the preparation is known as *syrup,* or *simple syrup* if the sucrose concentration is 85%. Syrups are also used to apply sugar coatings to tablets, particularly those with disagreeable aromas or acrid taste. In addition to sucrose, certain other polyols, such as glycerin or sorbitol, may be added to retard crystallization of sucrose or to increase the solubility of added ingredients. Alcohol often is included as a preservative and also as a solvent for flavors; further resistance to microbial attack can be enhanced by incorporating antimicrobial agents. When the aqueous preparation contains some added medicinal substance, the syrup is called a *medicated syrup. Flavored* syrups are usually not medicated, but rather contain various aromatic or pleasantly flavored substances and are intended to be used as a vehicle or flavor for prescriptions, such as Acacia, Cherry, Cocoa, Orange, and Raspberry USP.

Flavored syrups offer unusual opportunities as vehicles in extemporaneous compounding and are accepted readily by both children and adults. Because they contain no, or very little, alcohol they are vehicles of choice for many of the drugs that are prescribed by pediatricians. Their lack of alcohol makes them superior solvents for water-soluble substances. However, sucrose-based medicines continuously administered to children apparently cause an increase in dental caries and gingivitis; consequently, alternate formulations of the drug either unsweetened or sweetened with noncariogenic substances should be considered. A knowledge of the sugar content of liquid medicines is useful for patients who are on a restricted calorie intake; a list has been prepared by Greenwood.[49]

As noted above, sucrose-based syrups may be substituted in whole or in part by other agents in the preparation of medicated syrups. A solution of sorbitol, or a mixture of polyols, such as sorbitol and glycerin, is commonly used. Sorbitol Solution, USP, which contains 64% by weight of the polyhydric alcohol sorbitol, is often used in sugar-free and children's preparations. However, reports of adverse reactions to sorbitol are largely due to its action as an osmotic laxative when ingested orally.[50] Ingestion of large quantities of sorbitol (> 20 g/day in adults) should therefore be avoided.

Syrups possess remarkable taste-masking properties for bitter or saline drugs. Syrups flavored with Glycyrrhizin, a triterpene glycoside extracted from licorice root, has been recommended for disguising the salty taste of bromides, iodides, and chlorides.[51] This has been attributed to its colloidal character and its double sweetness—the immediate sweetness of the sugar and the lingering sweetness of the glycyrrhizin. This syrup is also of value in masking bitterness in preparations containing the B complex vitamins. Acacia Syrup USP is of particular value as a vehicle for masking the disagreeable taste of many medicaments because of its colloidal character. Raspberry Syrup, USP is one of the most efficient flavoring agents and is especially useful in masking the taste of bitter drugs. Many factors, however, enter into the choice of a suitable flavoring agent. Literature reports are often contradictory and there appears to be no substitute for the taste panel when developing new formulations.[52]

It is important that the concentration of sucrose approach but not quite reach the saturation point. In dilute solutions sucrose provides an excellent nutrient for molds, yeasts, and other microorganisms. In concentrations of 65% by weight or more, the solution will retard the growth of such microorganisms. However, a saturated solution may lead to crystallization of a part of the sucrose under conditions of changing temperature. Several commercial medicated syrups are available for a variety of indications (Table 39-5).

Preparation of Syrups

Syrups are generally prepared using one of four techniques: solution with heat, solution by agitation, addition of sucrose to a liquid medication or flavored liquid, and percolation. The method of choice depends on the physical and chemical characteristics of the substances entering into the preparation. In many cases, syrups may be successfully prepared by more than one of the above methods, and the selection may simply be a matter of preference on the part of the pharmacist. Many of the compendial syrups do not have a designated method for preparation because most are commercially available and are not prepared extemporaneously by the pharmacist.

Solution with Heat is a suitable preparation method if the constituents are not volatile or degraded by heat, and when it is desirable to make the syrup rapidly. Purified water is heated to 80°–85°C, removed from its heat source, and sucrose is added with vigorous agitation. Then, other required heat-stable components are added to the hot syrup, the mixture is allowed to cool, and its volume is adjusted to the proper level by the addition of purified water. In instances in which heat labile agents or volatile substances, such as flavors and alcohol, are to be

Table 39-5. Examples of Commercial Medicated Syrups

PRODUCT NAME	MANUFACTURER	ACTIVE INGREDIENT & DOSE	INDICATION
Chlor-Trimeton	Schering-Plough	2 mg chlorpheniramine maleate / 5 mL	Allergic rhinitus
Children's Benadryl	Pfizer	12.5 mg diphenhydramine HCl / 5 mL	Allergic rhinitus
Demerol Syrup	Sanofi	50 mg meperidine HCl / 5 mL	Narcotic analgesic
Ditropan Syrup	Ortho-McNeil	5 mg Oxybutynin chloride / 5 mL	Overactive bladder
Dramamine	Pfizer	12.5 mg dimenhydrinate / 5 mL	Antiemitic
Phenergan Syrup	Wyeth-Ayerst	25 mg promethazine HCl / 5 mL	Antiemitic
Symmetrel Syrup	Endo	50 mg amantadine HCl / 5 mL	Antiviral

added, they are generally incorporated into the syrup after cooling to room temperature.

When heat is used in the preparation of syrups, there is almost certain to be an inversion of a slight portion of the sucrose. Sucrose, a disaccharide, may be hydrolyzed into monosaccharides, dextrose (glucose), and fructose (levulose). This hydrolytic reaction is referred to as *inversion*, and the combination of the two monosaccharide products is *invert sugar*. Sucrose solutions are dextrorotary, but as hydrolysis proceeds, the optical rotation decreases and becomes negative when the reaction is complete. The rate of inversion is increased greatly by the presence of acids; the hydrogen ion acts as a catalyst in this hydrolytic reaction. Invert sugar is more readily fermentable than sucrose and tends to be darker in color. Nevertheless, its two reducing sugars are of value in retarding the oxidation of other substances.

The fructose formed during inversion is sweeter than sucrose, and thus the resulting syrup is sweeter than the original syrup. The relative sweetness of fructose, sucrose, and dextrose is in the ratio of 173:100:74. Thus, invert sugar is $1/100 (173 + 74)1/2 = 1.23$ times as sweet as sucrose. Fructose is responsible for the darkening of syrup, as it is amber in color. If the syrup is significantly overheated, sucrose is carmelized and becomes darker. Excessive heating of syrups is undesirable because inversion occurs with an increased tendency to ferment. Syrups cannot be sterilized in an autoclave without some caramelization.

Agitation without Heat is used in cases in which heat would cause degradation or volatilize formulation constituents. On a small scale, sucrose and other formulation ingredients may be dissolved in purified water by placing the ingredients in a vessel of greater capacity than the volume of syrup to be prepared, allowing intense agitation without spillage. This process is more time-consuming than solution with heat, but the product has greater stability. Large glass-lined and stainless steel tanks equipped with mechanical mixers are employed in the large scale preparation of syrups.

Often, simple syrup or some other non-medicated syrup, rather than sucrose, is employed as the sweetening agent and vehicle. When solid agents are to be added to a syrup, it is best to dissolve them in a minimal amount of purified water and then incorporate the resulting solution into the syrup. When solid substances are added directly to syrups, they dissolve slowly because the viscous nature of the syrup does not permit the solid substance to distribute readily.

This method and that previously described are used for the preparation of a wide variety of preparations that are described popularly as syrups. Most cough syrups, for example, contain sucrose and one or more active ingredients. Many other active ingredients (eg, ephedrine sulfate, dicyclomine hydrochloride, chloral hydrate, or chlorpromazine hydrochloride) are marketed as syrups. Like cough syrups, these preparations are flavored, colored, and recommended in those instances where the patient cannot swallow the solid dosage form.

Addition of sucrose to a liquid medication or flavored liquid is often used with fluidextracts, tinctures. Syrups made in this way usually develop precipitates because alcohol is often an ingredient of the liquids thus used, and the resinous and oily substances solubilized by the alcohol precipitate when water is added. A modification of this process entails mixing the fluidextract or tincture with the water, allowing the mixture to stand to permit the separation of insoluble constituents, filtering, and then dissolving the sucrose in the filtrate. It is obvious that this procedure is not permissible when the precipitated ingredients are the valuable medicinal agents.

In the *percolation* method, either purified water or the source of the medicinal component is passed slowly through a bed of crystalline sucrose, thus dissolving it and forming a syrup. This latter method really involves two separate procedures: first the preparation of the extractive of the drug and then the preparation of the syrup. To be successful in using this process, technique is critical: (1) the percolator used should be cylindrical or semicylindrical and cone-shaped as it nears the lower orifice; (2) a coarse granular sugar must be used, otherwise it will coalesce into a compact mass, which the liquid cannot permeate. The percolation method is applied on a commercial scale for the making of compendial syrups as well as those for confectionary use.

Ipecac syrup is prepared by percolation by adding glycerin and syrup to an extractive of powdered ipecac obtained by percolation. The drug ipecac consists of the dried rhizome and roots of *Cephaelis ipecacuanha* and contains the medicinally active alkaloids, emetine, cephaeline, and psychotrine. These alkaloids are extracted from the powdered ipecac by percolation with a hydroalcoholic solvent. The syrup is categorized as an emetic with a usual dose of 15 mL. This amount of syrup is commonly used in the management of poisoning in children when the evacuation of stomach contents is desirable. About 80% of children given this dose will vomit within a half hour. Bulimics have used ipecac to bring on attacks of vomiting in an attempt to lose more weight.[53] Pharmacists must be aware of this abuse and warn these individuals because one of the active ingredients is emetine. With chronic abuse of the syrup, emetine builds up toxic levels within body tissues and in 3 to 4 months can do irreversible damage to heart muscles resulting in symptoms mimicking a heart attack.

Syrups should be made in quantities that can be consumed within a few months, except in those cases where special facilities can be employed for their preservation; a low temperature is the best method. Concentration without super-saturation is also a condition favorable to preservation. The USP states that syrups may contain preservatives. Glycerin, methylparaben, benzoic acid, and sodium benzoate may be used to prevent bacterial and mold growth. Combinations of alkyl esters of *p*-hydroxybenzoic acid are effective inhibitors of yeasts that have been implicated in the contamination of commercial syrups. Syrups should be preserved in well-dried bottles, preferably those that have been sterilized. These bottles should not hold more than is likely to be required during 4 to 6 weeks and should be filled completely, carefully closed, and stored in a cool, dark place.

Some examples of syrup formulations are noted below:

Ferrous Sulfate Syrup

Ferrous Sulfate	40.0 g
Citric Acid	2.1 g
Peppermint Spirit	2 mL
Sucrose	825 g
Purified Water	to make 1000.0 mL

Dissolve the Ferrous Sulfate, Citric Acid, Peppermint Spirit, and 200 g of the Sucrose in 450 mL of Purified Water, and filter the solution until clear. Dissolve the remainder of the Sucrose in the clear filtrate, and add Purified Water to make 1000 mL. Mix, and filter, if necessary, through a pledget of cotton.

Amantadine Hydrochloride Syrup

Amantadine Hydrochloride	10.0 g
Citric Acid	2.1 g
Artifical Raspberry Flavor	2 mL
Methyl paraben	2 g
Propyl paraben	0.5 g
Sorbitol Solution	to make 1000 mL

Dissolve the amantadine hydrochloride, the Citric Acid, flavor and preservatives in the sorbitol solution.

Syrups are useful for preparing liquid oral dosage forms from not only the pure drug, as described above, but also injections, capsules, or tablets if the pure drug is not readily available. If the drug and all the excipients in the preparation, such as injectables or capsules, are water-soluble, a solution should result if a syrup is prepared. On the other hand, if the preparation to be used contains water-insoluble ingredients, as is usually the case with tablets and some capsules, a suspension will be formed. Several of these preparations have been described in

the literature, in regard to their formulation, stability, and bioavailability. Some drugs that have been prepared from either the pure drug or an injectable form include midazolam, atropine, aminocaproic acid, terbutaline, procainamide, chloroquine, propranolol, and citrated caffeine.[54,55] If the appropriate salt of the drug is used, a solution will result.

When tablets are introduced to a syrup formulation, a suspension is often formed because there are water-insoluble ingredients used in tablet preparations. Examples of medicated syrups prepared from tablets are clonidine hydrochloride, cefuroxime axetil, famotidine, terbutaline sulfate, spironolactone, ranitidine, and rifampin.[56,57] The resulting suspensions should have a uniform distribution of particles so that a consistent dose is obtained. If the materials are not distributed uniformly, more appropriate suspending formulations should be considered, which are described later in the chapter. If pharmaceutical preparations contain a liquid that is insoluble in water, such as valproic acid or simethicone, to be incorporated into syrups, an emulsion will form and it will be difficult to prepare a uniform product.

Honeys

Honeys are thick liquid preparations somewhat allied to the syrups, differing in that honey, instead of syrup, is used as a base. They are unimportant as a class of preparations today, but at one time, before sugar was available and honey was the most common sweetening agent, they were used widely. Honey and sugar pastes are used to a small extent and have been discussed in the pharmaceutical literature for topical application for the treatment of certain types of ulcers and abscesses.[58]

Mucilages

Mucilages are thick, viscid, adhesive liquids, produced by dispersing gum in water, or by extracting the mucilaginous principles from vegetable substances with water. The mucilages all are prone to decomposition, showing appreciable decrease in viscosity on storage; they should never be made in quantities larger than can be used immediately, unless a preservative is added. Mucilages are used primarily to aid in suspending insoluble substances in liquids; their colloidal character and viscosity help prevent immediate sedimentation. Examples include sulfur in lotions, resin in mixtures, and oils in emulsions. Both tragacanth and acacia either are partially or completely insoluble in alcohol. Tragacanth is precipitated from solution by alcohol, but acacia, on the other hand, is soluble in diluted alcoholic solutions. A 60% solution of acacia may be prepared with 20% alcohol, and a 4% solution of acacia may be prepared even with 50% alcohol.

Recent research on mucilages includes the preparation of mucilage from plantain and the identification of its sugars, the preparation and suspending properties of cocoa gum, the preparation of glycerin ointments using flaxseed mucilage, and the consideration of various gums and mucilages obtained from several Indian plants for pharmaceutical purposes.

Several synthetic mucilage-like substances such as *polyvinyl alcohol, methylcellulose, carboxymethylcellulose,* and related substances are used at the appropriate concentration as mucilage substitutes, and emulsifying and suspending agents. Methylcellulose is used widely as a bulk laxative because it absorbs water and swells to a hydrogel in the intestine, in much the same manner as *psyllium* or *karaya* gum. Methylcellulose Oral Solution USP is a flavored solution of the agent. It may be prepared by adding slowly the methylcellulose to about one-third the amount of boiling water, with stirring, until it is thoroughly wetted. Cold water then should be added and the wetted material allowed to dissolve while stirring. The viscosity of the solution will depend upon the concentration and the specifications of the methylcellulose. The synthetic gums are non-glycogenetic and may be used in the preparation of diabetic syrups. Sodium carboxymethyl cellulose of a medium grade in water (0.25–1%) is generally suitable for preparing a suspending vehicle. Several formulas for such syrups, based on sodium carboxymethylcellulose, have been proposed.

Uniformly smooth mucilages sometimes are difficult to prepare because of the uneven wetting of the gums. In general, it is best to use fine gum particles and disperse them with agitation in a small quantity of 95% alcohol or in cold water (except for methylcellulose). The appropriate amount of water then can be added with constant stirring. A review of the chemistry and properties of acacia and other gums has been prepared.[59]

Jellies

Jellies are a class of gels in which the structural coherent matrix contains a high portion of liquid, usually water. They are similar to mucilages, in that they may be prepared from similar gums, but they differ from the latter in having a jelly-like consistency. A whole gum of the best quality, rather than a powdered gum, is desirable to obtain a clear preparation of uniform consistency. Although the specific thickening agent in the USP jellies is not indicated, reference usually is made in the monograph to a water-soluble, sterile, viscous base. These preparations also may be formulated with water from acacia, chondrus, gelatin, carboxymethylcellulose, hydroxyethylcellulose, and similar substances.

Jellies are used as lubricants for surgical gloves, catheters, and rectal thermometers. Lidocaine Hydrochloride Jelly USP is used as a topical anesthetic. Therapeutic vaginal jellies are available and certain jelly-like preparations are used for contraceptive purposes, which often contain surface-active agents to enhance the spermatocidal properties of the jelly. Aromatics, such as methyl salicylate and eucalyptol, often are added to give the preparation a desirable odor.

Jellies are prone to microbial contamination and therefore contain preservatives; for example, methyl *p*-hydroxybenzoate is used as a preservative in a base for medicated jellies. One base contains sodium alginate, glycerin, calcium gluconate, and water. The calcium ions cause a cross-linking with sodium alginate to form a gel of firmer consistency. A discussion of gels is provided later in the chapter.

NONAQUEOUS SOLUTIONS

It is difficult to evaluate fairly the importance of nonaqueous solvents in pharmaceutical processes. That they are important in the manufacture of pharmaceuticals is an understatement. However, pharmaceutical preparations are, and, in particular, those intended for internal use, rarely contain more than minor quantities of the organic solvents that are common to the manufacturing or analytical operation. Products of commerce for internal use may contain solvents such as ethanol, glycerin, propylene glycol, certain oils, and liquid paraffin. Preparations intended for external use may contain solvents in addition to those just mentioned, namely isopropyl alcohol, polyethylene glycols, various ethers, and certain esters.

Although the lines between aqueous and nonaqueous preparations tend to blur in those cases where the solvent is water-soluble, it is possible to categorize a number of products as nonaqueous. This section is, therefore, devoted to groups of non-

aqueous solutions: the alcoholic or hydroalcoholic solutions (eg, elixirs and spirits), ethereal solutions (eg, collodions), glycerin solutions (eg, glycerins), oleaginous solutions (eg, liniments, oleovitamins, and toothache drops), inhalations, and inhalants.

Although the above list is limited, a wide variety of solvents are used in various pharmaceutical preparations. Solvents such as glycerol formal, dimethylacetamide, and glycerol dimethylketal have been suggested for some products produced by the industry. However, the toxicity of many of these solvents is not well established and, for this reason, careful clinical studies should be carried out on the formulated product before it is released to the marketplace. It is essential that the toxicity of solvents be tested appropriately and approved to avoid problems; for example, lives were lost in 1937 when diethylene glycol was used in an elixir of sulfanilamide. The result of this tragedy was the 1938 Federal Food, Drug, and Cosmetic Act, which required that products be tested for both safety and effectiveness.

COLLODIONS

Collodions are liquid preparations containing pyroxylin, a partially nitrated cellulose, in a mixture of ethyl ether and ethanol. They are applied to the skin by means of a soft brush or other suitable applicator and, when the ether and ethanol have evaporated, leave a film of pyroxylin on the surface. Salicylic Acid Collodion USP, contains 10% *w/v* of salicylic acid in Flexible Collodion USP and is used as a keratolytic agent in the treatment of corns and warts. Collodion USP and Flexible Collodion USP are water-repellent protectives for minor cuts, scratches, and chigger bites. Collodion is made flexible by the addition of castor oil and camphor. Collodion has been used to reduce or eliminate the side effects of fluorouracil treatment of solar keratoses.[60] Vehicles other than flexible collodion, such as a polyacrylic base, have been used to incorporate salicylic acid for the treatment of warts with less irritation.

ELIXIRS

Elixirs are clear, pleasantly flavored, sweetened hydroalcoholic liquids intended for oral use. The main ingredients in elixirs are ethanol and water but glycerin, sorbitol, propylene glycol, flavoring agents, preservatives, and syrups often are used in the preparation of the final product. The solvents are often used to increase the solubility of the drug substance in the dosage form. Elixirs are more fluid than syrups, due to the use of less viscous ingredients such as alcohol and the minimal use of viscosity-improving agents such as sucrose. They are used as flavors and vehicles such as Aromatic Elixir USP for drug substances; when such substances are incorporated into the specified solvents, they are classified as medicated elixirs, such as Dexamethasone Elixir USP and Phenobarbital Elixir USP.

The distinction between some of the medicated syrups and elixirs is not always clear. For example, Ephedrine Sulfate Syrup USP contains between 20 and 40 mL of alcohol in 1000 mL of product. Definitions are sometimes inconsistent and, in some instances, not too important with respect to the naming of the articles of commerce. To be designated as an elixir, however, the solution must contain alcohol. The alcoholic content will vary greatly, from elixirs containing only a small quantity to those that contain a considerable portion as a necessary aid to solubility. For example, Aromatic Elixir USP contains 21% to 23% alcohol; Compound Benzaldehyde Elixir USP, on the other hand, contains 3% to 5%.

Elixirs also may contain glycerin and syrup. These may be added to increase the solubility of the medicinal agent, for sweetening purposes, or to decrease the pharmacological effects of the alcohol. Some elixirs contain propylene glycol. Claims have been made for this solvent as a satisfactory substitute for both glycerin and alcohol.

Although alcohol is an excellent solvent for some drugs, it does accentuate the saline taste of bromides and similar salts. It often is desirable, therefore, to substitute some other solvent that is more effective in masking such tastes for part of the alcohol in the formula. In general, if taste is a consideration, the formulator is more prone to use a syrup rather than a hydroalcoholic vehicle.

Because only relatively small quantities of ingredients have to be dissolved, elixirs are more readily prepared and manufactured than syrups, which frequently contain considerable amounts of sugar. An elixir may contain both water- and alcohol-soluble ingredients. If such is the case, the following procedure is indicated:

> Dissolve the water-soluble ingredients in part of the water. Add and solubilize the sucrose in the aqueous solution. Prepare an alcoholic solution containing the other ingredients. Add the aqueous phase to the alcoholic solution, filter, and make to volume with water.

Sucrose increases viscosity and decreases the solubilizing properties of water and so must be added after the primary solution has been effected. A high alcoholic content is maintained during preparation by adding the aqueous phase to the alcoholic solution. Elixirs always should be brilliantly clear. They may be strained or filtered and, if necessary, subjected to the clarifying action of purified talc or siliceous earth.

Elixirs, and many other liquid preparations intended for internal use, such as the diabetic syrups thickened with sodium carboxymethylcellulose or similar substances, contain saccharin, aspartame, acesulfame potassium, and other sweeteners. Cyclamates and saccharin have been banned in some countries as ingredients in manufactured products. Much research has been done to find a safe synthetic substitute for sucrose.

Research concerning the preparation of a dry elixir has been conducted by Kim and co-workers.[61] Dry Elixirs containing a nonsteroidal anti-inflammatory drug and ethanol were encapsulated in a dextrin. The dissolution rate constant of the drug from the microcapsules usually increased considerably compared to the drug alone, possibly due to the cosolvent ethanol. It is suggested that this type of dosage form may be useful to improve the solubility, dissolution rate, and bioavailability of the drug.

Because elixirs contain alcohol, incompatibilities of this solvent are an important consideration during formulation. Alcohol precipitates tragacanth, acacia, and agar from aqueous solutions. Similarly, it will precipitate many inorganic salts from similar solutions. The implication here is that such substances should be absent from the aqueous phase or present in such concentrations that there is no danger of precipitation on standing.

If an aqueous solution is added to an elixir, a partial precipitation of alcohol soluble ingredients may occur. This is due to the reduced alcoholic content of the final preparation. Usually, however, the alcoholic content of the mixture is not sufficiently decreased to cause separation. As vehicles for tinctures and fluidextracts, the elixirs generally cause a separation of extractive matter from these products due to a reduction of the alcoholic content. Many of the incompatibilities between elixirs, and the substances combined with them, are due to the chemical characteristics of the elixir per se, or of the ingredients in the final preparation. Thus, certain elixirs are acid in reaction while others may be alkaline and will, therefore, behave accordingly.

Some example formulations of medicated elixirs are as follows:

Phenobarbital Elixir

Phenobarbital	4.00 g
Propylene Glycol	50 mL
Alcohol	200 mL
Sorbitol Solution	600 mL
Saccharin Sodium	5.0 g
Flavor	qs
Purified Water, to make	1000 mL

Theophylline Elixir

Theophylline	5.3 g
Citric Acid	10.0 g
Syrup	132.0 mL
Glycerin	50.0 mL
Sorbitol Solution	324.0 mL
Alcohol	200.00 mL
Flavor	q.s
Purified Water, to make	1000.0 mL

GLYCERINS

Glycerins or glycerites are solutions or mixtures of medicinal substances in not less than 50% by weight of glycerin. Most of the glycerins are extremely viscous and some are of a jelly-like consistency. Few of them are used extensively. Glycerin is a valuable pharmaceutical solvent forming permanent and concentrated solutions not otherwise obtainable. Glycerin is used as the sole solvent for the preparation of Antipyrine and Benzocaine Otic Solution USP. Glycerins are hygroscopic and should be stored in tightly closed containers.

INHALATIONS AND INHALANTS

Inhalation preparations are so used or designed that the drug is carried into the respiratory tree of the patient. The vapor or mist reaches the affected area and gives prompt relief from the symptoms of bronchial and nasal congestion. The USP defines Inhalations in the following way:

"Inhalations are drugs or solutions or suspensions of one or more drug substances administered to the nasal or oral respiratory route for local or systemic effect. Solutions of drug substances in sterile water for inhalation or in sodium chloride inhalation solution may be nebulized by the use of inert gases. Nebulizers are suitable for the administration of inhalation solutions only if they give droplets sufficiently fine and uniform in size so that the mist reaches the bronchioles. Nebulized solutions may be breathed directly from the nebulizer, or the nebulizer may be attached to a plastic face mask, tent or intermittent positive pressure breathing (IPPB) machine."

Another group of products, also known as metered-dose inhalers (MDIs) are propellant-driven drug suspensions or solutions in liquefied gas propellant (chlorofluorocarbons and hydrofluoroalkanes) with or without a cosolvent and are intended for delivering metered doses of the drug to the respiratory tract. An MDI contains multiple doses, often exceeding several hundred. The most common single-dose volumes delivered are from 25 to 100 μL (also expressed as mg) per actuation. Examples of MDIs containing drug solutions are Epinephrine Inhalation Aerosol, USP and Isoproterenol Hydrochloride and Phenylephrine Bitartrate Inhalation Aerosol, respectively. Both the solubility and stability of the drug in the propellant mixture must be investigated during formulation development. Ethanol is commonly used as a cosolvent hydrofluoroalkane propellants, and was reported to significantly increase the solubility of steroids.[62]

As stated in the USP, particle size is of major importance in the administration of this type of preparation. The various mechanical devices that are used in conjunction with inhalations are described in Chapter 50 (*Aerosols*). It has been reported that the optimum particle size for penetration into the pulmonary cavity is of the order of 0.5 to 7.0 μm.[63] Fine mists are produced by pressurized aerosols and hence possess basic advantages over the older nebulizers; in addition, metered aerosols deliver more uniform doses. A number of inhalations are described in the USP.

The USP defines "inhalants" as follows:

"A special class of inhalations termed "inhalants" consists of drugs or combinations of drugs that, by virtue of their high vapor pressure, can be carried by an air current into the nasal passage where they exert their effect. The container from which the inhalant is administered is known as an inhaler."

Amyl nitrate USP and Propylhexedrine Inhalant USP are two examples. Amyl nitrite is a clear, yellowish, volatile liquid that acts as a vasodilator when inhaled. The drug is prepared in sealed glass vials that are covered with a protective gauze cloth. Upon use, the glass vial is broken in the fingertips and the cloth soaks up the liquid which is then inhaled. The vials generally contain 0.3 mL of the drug substance. The effects of the drug are rapid and are used in the treatment of anginal pain.

Propylhexedrine is the active ingredient in the widely used Benzedrex Inhaler. Propylhexedrine is a liquid, vasoconstrictor agent that volatilizes slowly at room temperature. This quality enables it to be effectively used as an inhalant. The official inhalant consists of cylindrical rolls of suitable fibrous material impregnated with propylhexedrine, usually aromatized to mask its amine-like odor, and contained in a suitable inhaler. The vapor of the drug is inhaled into the nostrils when needed to relieve nasal congestion due to colds and hay fever. It may also be employed to relieve ear block and the pressure pain in air travelers. Each plastic tube of the commercial product contains 250 mg of propylhexedrine with aromatics. The containers should be tightly closed after each opening to prevent loss of the drug vapors.

LINIMENTS

Liniments are alcoholic or oil-based solutions or emulsions containing therapeutic agents intended for external application. These preparations may be liquids or semisolids that are rubbed onto the affected area; because of this, they were once called *embrocations*.

Liniments usually are applied with friction and rubbing of the skin, the oil or soap base providing for ease of application and massage. Alcoholic liniments are used generally for their rubefacient, counterirritant, mildly astringent, and penetrating effects. Such liniments penetrate the skin more readily than do those with an oil base. The oily liniments, therefore, are milder in their action but are more useful when massage is required. Depending on their ingredients, such liniments may function solely as protective coatings. Liniments should not be applied to skin that is bruised or broken.

Other liniments contain antipruritics, astringents, emollients, or analgesics and are classified on the basis of their active ingredient. Dermatologists prescribe products of this type but only those containing the rubefacients are advertised extensively and used by consumers for treating minor muscular aches and pains. It is essential that these applications be marked clearly "For External Use Only". Liniments containing a capsaicin are being investigated for treatment of pruritus.[64]

OLEOVITAMINS

Oleovitamins are fish liver oils diluted with edible vegetable oil or solutions of the indicated vitamins or vitamin concentrates (usually vitamin A and D) in fish liver oil. The definition is broad enough to include a wide variety of marketed products.

In oleovitamin A and D, USP, vitamin D may be present as ergocalciferol or cholecalciferol obtained by the activation of ergosterol or 7-dehydrocholesterol, or may be obtained from natural sources. Synthetic vitamin A, or a concentrate, may be used to prepare oleovitamin A. The starting material for the concentrate is fish liver oil, the active ingredient being isolated by molecular distillation or by a saponification and extraction procedure. These vitamins are unstable in the presence of rancid oils; therefore, these preparations should be stored in small, tight containers, preferably under vacuum or under an atmosphere of an inert gas, protected from light and air.

SPIRITS

Spirits, sometimes known as essences, are alcoholic or hydroalcoholic solutions of volatile substances. Like the aromatic waters, the active ingredient in the spirit may be a solid, liquid, or gas. The genealogical tree for this class of preparations begins with a distinguished pair of products, Brandy (*Spiritus Vini Vitis*) and Whisky (*Spiritus Frumenti*), and ends with a wide variety of products that comply with the definition given above. Physicians have debated the therapeutic value of the former products, and these are no longer compendial.

Generally, the alcohol concentration of spirits is rather high, usually over 60%. Because of the greater solubility of aromatic or volatile substances in alcohol than in water, spirits can contain a greater concentration of these materials than the corresponding aromatic waters. When mixed with water or with an aqueous preparation, the volatile substances present in spirits generally separate from solution and form a milky preparation. Salts may be precipitated from their aqueous solutions by the addition of spirits due to their lesser solubility in alcoholic liquids. Some spirits show incompatibilities characteristic of the ingredients they contain. For example, Aromatic Ammonia Spirit cannot be mixed with aqueous preparations containing alkaloids (eg, codeine phosphate). An acid–base reaction (ammonia-phosphate) occurs, and if the alcohol content of the final mixture is too low, codeine will precipitate. Spirits should be stored in tight, light-resistant containers and in a cool place. This tends to prevent evaporation and volatilization of either the alcohol or the active principle and to limit oxidative changes.

Spirits may be used pharmaceutically as flavoring agents and medicinally for the therapeutic value of the aromatic solute. As flavoring agents they are used to impart the flavor of their solute to other pharmaceutical preparations. For medicinal purposes, spirits may be taken orally, applied externally, or used by inhalation, depending upon the particular preparation. When taken orally, they are generally mixed with a portion of water to reduce the pungency of the spirit. Depending on the materials utilized, spirits may be prepared by simple solution, solution by maceration, or distillation. The spirits still listed in the USP/NF are aromatic ammonia spirit, camphor spirit, compound orange spirit, and peppermint spirit.

EMULSIONS

An emulsion is a two-phase system prepared by combining two immiscible liquids, in which small globules of one liquid are dispersed uniformly throughout the other liquid. The liquid that is dispersed into small droplets is called the dispersed, internal, or discontinuous phase. The other liquid is the dispersion medium, external phase, or continuous phase. Where oil is the dispersed phase and an aqueous solution is the continuous phase, the system is designated as an oil-in-water (O/W) emulsion. Conversely, where water or an aqueous solution is the dispersed phase and oil or oleaginous material is the continuous phase, the system is designated as a water-in-oil (W/O) emulsion. Emulsions may be employed orally, topically, or parenterally depending upon on the formulation ingredients and the intended application. Many pharmaceutical emulsions may not be classified as such because they are described by another pharmaceutical category more appropriately. For instance, certain lotions, liniments, creams, ointments, and commercial vitamin drops may be emulsions but may be preferentially referred to in these terms.

Emulsions possess a number of important advantages over other liquid forms:

- In an emulsion, poorly water-soluble drugs may be easily incorporated with improved dissolution rates and bioavailability.
- The unpleasant taste or odor of oils can be masked partially or wholly, by emulsification.
- The absorption rate and permeation of medicaments are can be controlled.
- Absorption may be enhanced by the diminished size of the internal phase.
- Formulation and technology for organ targeted delivery is available.
- Various particle sizes of the internal phase can be achieved by preparation technique, from micro emulsions (micron-sized particles) to nanoparticles.
- Water is an inexpensive diluent and a good solvent for the many drugs and flavors that are incorporated into an emulsion.

It is possible to prepare emulsions that are basically nonaqueous. For example, investigations of the emulsifying effects of anionic and cationic surfactants on the nonaqueous immiscible system, glycerin and olive oil, have shown that certain amines and three cationic agents produced stable emulsions. Although the USP definition is broad enough to encompass nonaqueous systems, emphasis is placed on those emulsions that contain water, as they are by far the most common in pharmacy.

When it is necessary to administer oils by the oral route, patient acceptance is enhanced when the oil is prepared in emulsion form. Thus, mineral oil (a laxative), valproic acid (an anticonvulsant), oil-soluble vitamins, vegetable oils, and preparations for enteral feeding are formulated frequently in an O/W emulsion form to enhance their palatability.

The bioavailability of oils for absorption may be enhanced when the oil is in the form of small droplets. Furthermore, the absorption of some drugs, such as griseofulvin may be enhanced when they are prepared in the form of an O/W emulsion.[65] Emulsion formulations of drugs such as erythromycin and physostigmine salicylate have been considered, in order to improve their stability.[66,67] Finally, the greatest use of emulsions is for topical preparations. Both O/W and W/O emulsions are used widely, depending upon the effect desired. Emulsion bases of the W/O type tend to be more occlusive and emollient than O/W emulsion bases, which tend to be removed more easily by water. The effects of viscosity, surface tension, solubility, particle size, complexation, and excipients on the bioavailability of emulsions have been reported.[68]

Although this section on emulsions focuses primarily on those for oral use and to a lesser degree those for topical application, it should be noted that there are a number of emulsions used parenterally that are described in specialized textbooks on this topic. For example, emulsions of the O/W type are used for intravenous feeding of lipid nutrients. These are used to provide a source of calories and essential fatty acids. These emulsions must meet exacting standards in regard to particle size, safety, and stability. Examples of commercial products include Diprivan Injectable Emulsion (*AstraZeneca*), EMLA Cream (*AstraZeneca*), Renova 0.02% Cream (*OrthoNeutrogena*), Bactroban Cream (*GlaxoSmithKline*), Cordran Lotion (*Watson*), Differin Cream (*Galderma*) and Renova 0.05% Cream (*OrthoNeutrogena*). Other specialized uses of emulsions include radiopaque emulsions that are used as diagnostic agents for x-ray examination.

THEORIES OF EMULSIFICATION

Several theories have been proposed to explain how emulsifying agents act in producing the multi-phase dispersion and in maintaining the stability of the resulting emulsion. Some of these theories apply to specific types of emulsifying agents and to certain conditions, such as pH of the system and the physicochemical nature and proportions of the internal and external phases). The most prevalent theories are the *surface-tension theory*, the *oriented-wedge theory*, and the *interfacial film theory*.

Liquids assume a shape to minimize their surface area, which is spherical for a small drop. In a spherical drop of liquid, there are attractive forces between the molecules, resisting distortion into a less spherical form. If two or more drops of the same liquid come into contact with one another, it is more thermodynamically favorable for them to coalesce, making a larger drop with a decreased surface area compared to the total surface area of the individual drops. The tendency of liquids to minimize their surface area can be measured quantitatively, and when the liquid is surrounded by air, the measurement is called the surface tension.

When a liquid is in contact with another liquid in which it is insoluble and immiscible, the force causing each liquid to resist breaking up into smaller particles is called interfacial tension. Surface active agents, or surfactants, are substances that reduce the resistance of a droplet to form smaller droplets. Surfactants are also called emulsifiers and wetting agents. According to the surface tension theory of emulsification, the use of surfactants results in a reduction in the interfacial tension of the two immiscible liquids, reducing the repellent force between the liquids and diminishing each liquid's attraction for its own molecules. Thus, surfactants enable large globules to break into smaller ones, and prevent small globules from coalescing into larger ones.

The oriented wedge theory proposes that the surfactant forms monomolecular layers around the droplets of the internal phase of the emulsion. The theory is based on the assumption that emulsifying agents orient themselves about and within a liquid relative to their solubility in that particular liquid. In a system containing two immiscible liquids, the emulsifying agent is preferentially soluble in one of the two liquids and becomes more embedded with that phase relative to the other. Many surfactants have a hydrophilic or water loving portion and a hydrophobic or water hating portion (but usually lipophilic or oil-loving), and the molecules will position or orient themselves into each phase. Depending upon the shape and size of the molecules, their solubility characteristics, and thus their orientation, the wedge shape theory proposes that emulsifiers will surround either oil globules or water globules.

Generally an emulsifying agent having a greater hydrophilic character than hydrophobic character will promote oil in water emulsions. On the other hand, water in oil emulsions result with the use of an emulsifyer that is more hydrophobic than hydrophilic. Putting it another way, the phase in which the emulsifying agent is more soluble will become the continuous or external phase of the emulsion. Although this theory does not represent a completely accurate depiction of the molecular arrangement of the emulsifier molecules, the concept that water soluble emulsifiers generally form oil in water emulsions is important.

The interfacial film theory proposes that the emulsifier forms an interface between the oil and water, surrounding the droplets of the internal phase as a thin layer of film adsorbed on the surface of the drops. The film prevents the contact and coalescing of the dispersed phase; the tougher and more pliable the film, the greater the stability of the emulsion. Naturally, the surfactant must be available to coat the entire surface of each drop of internal phase. Similar to the oriented wedge theory, the formation of an oil in water or a water in oil emulsion depends upon the degree of solubility of the emulsifier in the two phases, with water soluble agents encouraging oil in water emulsions and oil-soluble emulsifiers promoting water in oil emulsions.

In reality, none of the emulsion theories can individually explain the mechanism by which the many and varied emulsifiers promote emulsion formation and stability. It is more than likely that even within a given emulsion system, more than one of the theories of emulsification are applicable. For instance, reducing the interfacial tension is critical during initial formation of an emulsion, but the formation of a protective wedge of molecules or film of emulsifier is equally important for continued emulsion stability. Undoubtedly, many emulsifiers are capable of both tasks.

EMULSION FORMULATION INGREDIENTS

The first step in preparation of an emulsion is the selection of the emulsifier. The emulsifier must be compatible with the formulation ingredients and the active pharmaceutical ingredient. It should be stable, nontoxic, and promote emulsification to maintain the stability of the emulsion for the intended shelf life of the product. The selection of the oil phase for oral preparations depends upon the purpose of the product. For example, mineral oil is used as a laxative and corn oil is used for its nutrient properties. Vegetable oils can be used to dissolve or suspend pharmaceuticals such as oil-soluble vitamins.

Emulsions are thermodynamically unstable because of the large increase in surface energy that results from the combination of interfacial tension and large surface area of the dispersed phase and the different densities of the two phases. Thus, emulsions tend to cream—the less dense phase rises and the more dense phase falls in the container. Subsequently, the droplets can coalesce with a considerable reduction in surface free energy. Consequently, considerable research has been conducted on their preparation and stabilization. To prepare suitable emulsions that remain stable, a number of excipients are used in their preparation.

Emulsifiers often have a hydrophilic portion and a lipophilic portion with one or the other being more or less predominant. Griffin devised a method whereby emulsifying or surface-active agents may be categorized on the basis of their *hydrophilic-lipophilic balance or HLB value*. By this method, each agent is assigned an HLB value or number which is indicative of the substance's polarity, which may vary from 40 for sodium lauryl sulfate to 1 for oleic acid. Although the numbers have been assigned up to about 40, the usual range is between 1 and 20. Examples of HLB values for common emulsifiers used in pharmaceutical applications are listed in Table 39-6. HLB values have also been useful in describing the functional properties of materials. For example, HLB values from 1 to 3 typically exhibit anti foaming properties, values from 7 to 10 exhibit good wetting properties, values from 13 to 20 act as solubilizers, and values from 13 to 15 function as

Table 39-6. HLB Values of Common Emulsifiers Used in Pharmaceutical Systems

AGENT	HLB	CLASS
Oleic Acid	1.0	Anionic
Ethylene glycol distearate	1.5	Nonionic
Sorbitan tristearate (Span 65)	2.1	Nonionic
Glyceryl monooleate	3.3	Nonionic
Propylene glycol monostearate	3.4	Nonionic
Glyceryl monostearate	3.8	Nonionic
Sorbitan monooleate (Span 80)	4.3	Nonionic
Sorbitan monostearate (Span 60)	4.7	Nonionic
Diethylene glycol monolaurate	6.1	Nonionic
Sorbitan monopalmitate (Span 40)	6.7	Nonionic
Acacia	8.0	Anionic
Polyoxyethylene lauryl ether (Brij 30)	9.7	Nonionic
Polyoxyethylene monostearate (Myrj 45)	11.1	Nonionic
Triethanolamine oleate	12.0	Anionic
Polyoxyethylene sorbitan monostearate (Tween 60)	14.9	Nonionic
Polyoxyethylene sorbitan monooleate (Tween 80)	15.0	Nonionic
Polyoxyethylene sorbitan monolaurate (Tween 20)	16.7	Nonionic
Pluronic F 68	17.0	Nonionic
Sodium oleate	18.0	Anionic
Potassium oleate	20.0	Anionic
Cetrimonium Bromide	23.3	Cationic
Cetylpyridinium chloride	26.0	Cationic
Poloxamer 188	29.0	Nonionic
Sodium lauryl sulfate	40.0	Anionic

detergents. Oil in water emulsions typically have a weighted HLB value ranging from 8 to 16 while water in oil emulsions have weighted HLB values ranging from 3 to 8.

Materials that are highly polar or hydrophilic have been assigned higher numbers than materials that are less polar and more lipophilic. Generally, lipophilic surfactants have an HLB value from 0 to 10 and are known for their antifoaming, water in oil emulsifying or wetting properties. Hydrophilic surfactants have HLB values ranging from 10 to 20 and form oil in water emulsions. The HLB system also assigns values to oils and oil like substances. In using the HLB concept in the preparation of an emulsion, one selects emulsifying agents having the same or nearly the same HLB value as the oleaginous phase of the intended emulsion. When needed, two or more emulsifiers may be combined to achieve the proper HLB value.

The ionic nature of a surfactant is an important consideration when selecting a surfactant for an emulsion. Nonionic surfactants are effective over pH range 3 to 10; cationic surfactants are effective over pH range 3 to 7; and, anionic surfactants require a pH of greater than 8.[69]

Emulsifying agents may be divided into three classes: *natural emulsifying agents*, *finely divided solids,* and *synthetic emulsifying agents*.

1. **Natural Emulsifying Agents** are substances derived from vegetable sources and include acacia, tragacanth, alginates, chondrus, xanthan, and pectin. These materials form hydrophilic colloids when added to water and generally produce o/w emulsions. Although their surface activity is low, these materials achieve their emulsifying power by increasing the viscosity of the aqueous phase. Examples of emulsifying agents derived from animal sources include gelatin, egg yolk, casein, wool fat, cholesterol, wax, and lecithin. Because of the widely different chemical constitution of these compounds, they have a variety of uses, depending upon the specific compound, in both oral and topical preparations. All naturally occurring agents show variations in their emulsifying properties from batch to batch.

2. **Finely Divided Solids** are the colloidal clays: bentonite (aluminum silicate) and Veegum (magnesium aluminum silicate). These compounds are good emulsifiers and tend to be absorbed at the interface, increase the viscosity in the aqueous phase, and are often used in conjunction with a surfactant to prepare O/W emulsions. However, both O/W and W/O preparations can be prepared by adding the clay to the external phase. They are used frequently for external purposes such as a lotion or cream.

3. **Synthetic Emulsifying Agents** are very effective at lowering the interfacial tension between the oil and water phases because the molecules possess both hydrophilic and hydrophobic properties. These emulsifying agents are available in different ionic types: anionic, such as sodium dodecyl sulfate; cationic, such as benzalkonium chloride; nonionic, such as polyethylene glycol 400 monostearate; and ampholytic, such as long-chain amino acid derivatives. In addition to the emulsifying agents, *viscosity agents* are employed, namely the hydrophilic colloids such as naturally occurring gums, noted above, and partially synthetic polymers such as cellulose derivatives (eg, methylcellulose, hydroxypropylmethylcellulose, sodium carboxymethylcellulose) or a number of synthetic polymers that may be used, such as carbomer polymers. These materials are hydrophilic in nature and dissolve or disperse in water to give viscous solutions and function as emulsion stabilizers.

Other functional excipients are often utilized in emulsions. High molecular weight alcohols such as stearyl alcohol, cetyl alcohol, and glyceryl monostearate are employed primarily as thickening agents and stabilizers for o/w emulsions of certain lotions and ointments used externally. Cholesterol and cholesterol derivatives may also be employed in externally used emulsions and to promote w/o emulsions.

The aqueous phase of the emulsion favors the growth of microorganisms; because of this, a preservative usually is added to the product. Some of the preservatives that have been used include chlorocresol, chlorobutanol, mercurial preparations, salicylic acid, the esters of *p*-hydroxybenzoic acid, benzoic acid, sodium benzoate, or sorbic acid. The preservative should be selected with regard for the ultimate use of the preparation and possible incompatibilities between the preservative and the ingredients in the emulsion (eg, binding between the surfactant and the preservative). Low pH values of 5 to 6 and low concentrations of water are characteristics also likely to inhibit microbiological growth in emulsions.

Emulsions consist of an oil or lipid phase and an aqueous phase, thus the preservative may diffuse from the aqueous phase into the oil phase. It is in the aqueous phase that microorganisms tend to grow. As a result, water-soluble preservatives are more effective because the concentration of the unbound preservative in the aqueous phase assumes a great deal of importance in inhibiting the microbial growth. Esters of *p*-hydroxybenzoic acid appear to be the most satisfactory preservatives for emulsions.

Many mathematical models have been used to determine the availability of preservatives in emulsified systems. One model takes into account the O/W partition coefficient of the preservative, interaction of the preservative with the surfactant, interfacial tension and membrane permeability. However, because of the number of factors that reduce the effectiveness of the preservative, a final microbiological evaluation of the emulsion must be performed.

While emphasis concerning preservation of emulsions deals with the aqueous phase, microorganisms can reside also in the lipid phase. Consequently, it has been recommended that pairs of preservatives be used to ensure adequate concentration in both phases. Esters of *p*-hydroxybenzoic acid can be used to ensure appropriate concentrations in both phases because of their difference in oil and water solubilities.

The oxidative decomposition of certain excipients, the oil phase, and some pharmaceuticals is possible in emulsions, not only because of the usual amount of air dissolved in the liquid and the possible incorporation of air during the preparation of the product, but also the large interfacial area between the oil and water phase. The selection of the appropriate antioxidant briefly described at the beginning of the chapter depends on factors such as stability, compatibility with the ingredients of the emulsion, toxicity, effectiveness in emulsions, odor, taste, and distribution between the two phases.

PREPARATION OF EMULSIONS

After the purpose of the emulsions has been determined (eg, oral or topical use), the type of emulsions (O/W or W/O) and appropriate ingredients selected, and the theory of emulsification considered, then experimental formulations may be prepared. One method is suggested by Griffin[70]:

1. Group the ingredients on the basis of their solubilities in the aqueous and nonaqueous phases.
2. Determine the type of emulsion required and calculate an approximate HLB value.
3. Blend a low HLB emulsifier and a high HLB emulsifier to the calculated value. For experimental formulations, use a higher concentration of emulsifier (eg, 10% to 30% of the oil phase) than that required to produce a satisfactory product. Emulsifiers should, in general, be stable chemically, nontoxic, and suitably low in color, odor, and taste. The emulsifier is selected on the basis of these characteristics, as well as the type of equipment being used to blend the ingredients and the stability characteristics of the final product. Emulsions should not coalesce at room temperature, or when frozen and thawed repeatedly, or at elevated temperatures of up to 50°C. Mechanical energy input varies with the type of equipment used to prepare the emulsion. The more the energy input, the less the demand on the emulsifier. Both process and formulation variables can affect the stability of an emulsion.
4. Dissolve the oil-soluble ingredients and the emulsifiers in the oil. Heat, if necessary, to approximately 5° to 10°C over the melting point of the highest melting ingredient or to a maximum temperature of 70° to 80°C.
5. Dissolve the water-soluble ingredients (except acids and salts) in a sufficient quantity of water.
6. Heat the aqueous phase to a temperature that is 3° to 5°C higher than that of the oil phase.

7. Add the aqueous phase to the oily phase with suitable agitation.
8. If acids or salts are employed, dissolve them in water and add the solution to the cold emulsion.
9. Examine the emulsion and make adjustments in the formulation if the product is unstable. It may be necessary to add more emulsifier, to change to an emulsifier with a slightly higher or lower HLB value, or to use an emulsifier with different chemical characteristics.

The technique of emulsification of pharmaceutical preparations has been described by Nielloud and Marti-Mestres.[71] The preparation of an emulsion requires work to reduce the internal phase into small droplets and disperse them throughout the external phase. This can be accomplished by a mortar and pestle or a high-speed emulsifier. The addition of emulsifying agents not only reduces this work but also stabilizes the final emulsion.

Emulsions are prepared by four principal methods.

ADDITION OF INTERNAL PHASE TO EXTERNAL PHASE—This is usually the most satisfactory method for preparing emulsions as there is always an excess of the external phase present that promotes the type of emulsion desired. If the external phase is water and the internal phase is oil, the water-soluble substances are dissolved in the water and the oil-soluble substances mixed thoroughly in the oil. The oil mixture is added in portions to the aqueous preparation with agitation. To give a better shearing action during the preparation, sometimes all of the water is not mixed with the emulsifying agent until the primary emulsion with the oil is formed; subsequently, the remainder of the water is added. An example using gelatin Type A is given below.

ADDITION OF THE EXTERNAL PHASE TO THE INTERNAL PHASE, THE DRY GUM TECHNIQUE—Using an O/W emulsion as an example, the addition of the water (external phase) to the oil (internal phase) will promote the formation of a W/O emulsion due to the preponderance of the oil phase. After further addition of the water, phase inversion to an O/W emulsion should take place. This method is especially useful and successful when hydrophilic agents such as acacia, tragacanth, or methylcellulose are first mixed with the oil, effecting dispersion without wetting. Water is added, and eventually an O/W emulsion is formed. This "dry gum" technique is a rapid method for preparing small quantities of emulsion. The ratio 4 parts of oil, 2 parts of water and 1 part of gum provides maximum shearing action on the oil globules in the mortar. The emulsion then can be diluted and triturated with water to the appropriate concentration. The preparation of Mineral Oil Emulsion described below is an example.

MIXING BOTH PHASES AFTER HEATING—This method is used when waxes or other substances that require melting are used. The oil-soluble emulsifying agents, oils, and waxes are melted and mixed thoroughly. The water-soluble ingredients dissolved in the water are warmed to a temperature slightly higher than the oil phase. The two phases then are mixed and stirred until cold. For convenience, but not necessity, the aqueous solution is added to the oil mixture. This method frequently is used in the preparation of ointments and creams. An example of an oral preparation containing a poorly soluble drug is given below.

ALTERNATE ADDITION OF THE TWO PHASES TO THE EMULSIFYING AGENT—A portion of the oil, if an O/W emulsion is being prepared, is added to all of the oil-soluble emulsifying agents with mixing, then an equal quantity of water containing all the water-soluble emulsifying agents is added with stirring until the emulsion is formed. Further portions of the oil and water are added alternately until the final product is formed. The high concentration of the emulsifying agent in the original emulsion makes the initial emulsification more likely and the high viscosity provides effective shearing action leading to small droplets in the emulsion. This method often is used successfully with soaps.

Examples of some emulsions are given below.

Type A gelatin is prepared by acid-treated precursors and is used at a pH of about 3.2. It is incompatible with anionic emul-

sifying agents such as the vegetable gums. The following formula was recommended.

Type A Gelatin Emulsion

Gelatin (Type A)	8 g
Tartaric Acid	0.6 g
Flavor	q.s
Alcohol	60 mL
Oil	500 mL
Purified Water,	to make 1000 mL

Add the gelatin and the tartaric acid to about 300 mL of purified water, allow to stand for a few minutes, heat until the gelatin is dissolved, then raise the temperature to about 98°C and maintain this temperature for about 20 min. Cool to 50°C, add the flavor, the alcohol, and sufficient purified water to make 500 mL. Add the oil, agitate the mixture thoroughly, and pass it through a homogenizer or a colloid mill until the oil is dispersed completely and uniformly. This emulsion cannot be prepared by trituration or by the use of the usual stirring devices.

Type B gelatin is prepared from alkali-treated precursors and is used at a pH of about 8. It may be used with other anionic emulsifying agents but is incompatible with cationic types. If the emulsion contains 50% oil, 5 g of Type B gelatin, 2.5 g of sodium bicarbonate and sufficient tragacanth or agar should be incorporated into the aqueous phase to yield 1,000 mL of product of the required viscosity.

An emulsion that may be prepared by the mortar and pestle method is the following Mineral Oil Emulsion USP.

Mineral Oil Emulsion, USP

Mineral Oil	500 mL
Acacia, in very fine powder	125 g
Syrup	100 mL
Vanillin	40 mg
Alcohol	60 mL
Purified Water,	to make 1000 mL

The mineral oil and acacia are mixed in a dry Wedgwood mortar. Purified water (250 mL) is added and the mixture triturated vigorously until an emulsion is formed. A mixture of the syrup, 50 mL of purified water, and the vanillin dissolved in alcohol is added in divided portions with trituration; sufficient purified water is then added to the proper volume; the mixture is mixed well and homogenized.

An Oral Emulsion (O/W) Containing an Insoluble Drug[72]

Cottonseed Oil	460 g
Sulfadiazine	200 g
Sorbitan Monostearate	84 g
Polyoxyethylene 20	36 g
Sorbitan Monostearate	
Sodium Benzoate	2 g
Sweetener	qs
Flavor Oil	qs
Purified Water	1000 g

Heat the first three ingredients to 50°C and pass through colloid mill. Add the next four ingredients at 50°C to the first three ingredients at 65°C and stir while cooling to 45°C. Add the flavor oil and continue to stir until room temperature is reached.

PROPERTIES AND STABILITY OF EMULSIONS

The type of emulsion (O/W or W/O) depends, to some extent, on the phase to volume ratio. The higher the fraction of one phase, the greater likelihood it will form the external phase. Thus, O/W emulsions are favored if water forms a greater fraction of the volume than the oil phase. However, it is possible for the internal phase of an emulsion to occupy up to 74% of the volume of the emulsion and still form a stable product.

The consistency of emulsions can be increased by increasing the viscosity of the continuous phase, increasing the fractional volume of the internal phase, reducing the particle size of the internal phase, increasing the proportion of the emulsifying agent, or adding hydrophobic emulsifying agents to the oil phase of the emulsion.

The physical stability of emulsions may be defined by a number of expressions. The first of these, which is called *creaming,* is the movement of the droplets either upward or downward, depending upon their density. This gives a product that is not homogenous and can lead to poor content uniformity. Generally, creaming is not a serious problem because a moderate amount of shaking will redisperse the droplets uniformly. The rate of creaming may be decreased by considering the theory of creaming using Stokes' law. This equation relates the rate of creaming to the size of the droplets, the difference in densities, and the viscosity of the external phase. Thus, the rate of creaming may be decreased by decreasing the size of the droplets and increasing the viscosity of the external phases, both of which were discussed above. Minimizing the difference between densities is more challenging due to a number of practical difficulties.

When the droplets aggregate, they come together and act as a single unit, but do not fuse. As a result of the larger size, they tend to cream faster and further provoke physical instability. Aggregation is to some extent reversible and may be controlled by choosing a somewhat different surfactant system and controlling the electrical potential of the droplets. Coalescence of an emulsion is the fusion of the droplets, leading to a decrease in their numbers and eventually the complete separation of the two phases, yielding an unsatisfactory product that should be reformulated.

General methods are available for testing and challenging the stability of emulsions including bulk changes, centrifugal and ultracentrifugal studies, dielectric measurement, surface area measurement, temperature cycling, preservative effectiveness, and accelerated motion studies. Low-shear rheological studies measuring viscoelasticity are suggested as the optimal method of stability testing.

MULTIPLE EMULSIONS

A recent innovation in emulsion technology is the development of multiple emulsions. The dispersed phase of these emulsions contains even smaller droplets that are miscible with the continuous phase. Thus, the multiple emulsion may be O/W/O, where the aqueous phase is between two oil phases, or W/O/W, where the internal and external aqueous phases are separated by an oil phase. In these systems both hydrophobic and hydrophilic emulsifiers are used, and both have an effect on the yield and stability.[73,74]

It appears that O/W/O emulsions are formed better by lipophilic, nonionic surfactants using acacia emulsified simple systems, whereas W/O/W multiple emulsions are formed better by nonionic surfactants in a two-stage emulsification procedure. A specific formulation for a W/O/W emulsion may be prepared by forming the primary (W/O) emulsion from isopropyl myristate (47.5%), sorbitan monooleate (2.5%), and distilled water to 100%. This primary emulsion (50%) is added to a polyoxyethylene sorbitan monooleate (2% *w/v*) solution in water.[74] Other formulations of multiple emulsions include carboxymethylcellulose sodium, microcrystalline cellulose, sorbitan monooleate, and sorbitan trioleate.

Although the technique of preparing these emulsions is more complicated, research indicates potential use of these emulsions for prolonged action, taste-masking, more effective dosage forms, improved stability, parenteral preparations, protection against the external environment, and enzyme entrapment. These emulsions also may be used to separate two incompatible hydrophilic substances in the inner and outer aqueous phases by the middle oil phase. Some drugs that have been investigated in these types of emulsions are vancomycin[75] and prednisolone.[124]

MICROEMULSIONS

The coarse pharmaceutical macroemulsions appear white and tend to separate on standing. Microemulsions are translucent or transparent, do not separate, and typically have a droplet diameter from 10 to 200 nanometers. The microemulsions are not always distinguishable from micellar solutions. Both O/W and W/O emulsions are possible, and one may be converted to the other by addition of more internal phase or by altering the type of emulsifier. As the internal phase is added, the emulsion will pass through a viscoelastic gel stage; with further addition, an emulsion of the opposite type will occur.

The most obvious benefit of microemulsions is their stability. Usually, the emulsifier should be 20% to 30% of the weight of the oil used. The W/O systems are prepared by blending the oil and emulsifier with a little heat, if required, and then adding the water. The order of mixing for O/W systems is more flexible. One of the simplest methods is to blend the oil and the emulsifier and pour this into water with a little stirring.

If the emulsifier has been selected properly, microemulsification will occur almost spontaneously, leading to a satisfactory and stable preparation. The details of various preparations and the relationship between microemulsions and micellar solutions have been reviewed by Bourrel.[77] Other authors suggest that the preparation of microemulsions is considerably more difficult than the preparation of coarse emulsions. Rosano and colleagues discuss the use of a primary surfactant adsorbed at the interface that influences the curvature of the dispersed phase.[78] The amount of surfactant required may be estimated from the surface area of the droplets and the cross-sectional area of the surfactant molecule. The authors propose the use of a co-surfactant to form a duplex film and the order of mixing is important.

PROCESSING EQUIPMENT FOR EMULSIONS

The preparation of emulsions requires a certain amount of energy to form the interface between the two phases, and additional work must be done to stir the system to overcome resistance to flow. In addition, heat often is supplied to the system to melt waxy solids and/or reduce viscosity. Consequently, the preparation of emulsions on a large scale requires considerable amounts of energy for heating and mixing. Careful consideration of these processes has led to the development of low energy emulsification equipment that uses an appropriate emulsification temperature and selective heating of the ingredients. This process, described by Lin, involves the preparation of an emulsion concentrate subsequently diluted with the external phase at room temperature.[79,80]

Because of the variety of oils used, emulsifier agents, phase to volume ratios, and the desired physical properties of the product, a wide selection of equipment is available for preparing emulsions, and the main classes of equipment are discussed below. Homogenization speed and time and rate of cooling may influence the viscosity of the product. Further information may be obtained from the *Bibliography*.

Special techniques and equipment in certain instances will produce superior emulsions, including rapid cooling, reduction in particle size, or ultrasonic devices. A wide selection of equipment for processing both emulsions and suspensions has been described.[71] A number of improvements have been made to make the various processes more effective and energy efficient.

The mortar and pestle may be used to prepare small quantities of an emulsion in the pharmacy or laboratory, and it is one of the simplest and least expensive methods. It may be used for most of the different techniques of preparing emulsions. Generally, the final particle size is considerably larger than is achieved by the equipment described below. In addition, it is necessary for the ingredients to have a certain viscosity prior to trituration to achieve a satisfactory shear. Satisfactory

emulsions of low viscosity ingredients and small volumes may be prepared using the appropriate equipment described below.

Agitators

Ordinary agitation or shaking may be used to prepare the emulsion. This method frequently is employed by the pharmacist, particularly in the emulsification of easily dispersed, low-viscosity oils. Under certain conditions, intermittent shaking is considerably more effective than ordinary continuous shaking. Continuous shaking tends to break up not only the phase to be dispersed but also the dispersion medium, thus impairing the ease of emulsification. Laboratory shaking devices may be used for small-scale production.

Mechanical Mixers

Emulsions may be prepared by using one of several mixers that are available. Propeller and impeller type mixers that have a propeller attached to a shaft driven by an electric motor are convenient and portable and can be used for both stirring and emulsification. This type operates best in mixtures that have low viscosity, that is, mixtures with a viscosity of glycerin or less. They are also useful for preparing emulsions. A turbine mixer has a number of blades that may be straight or curved, with or without a pitch, mounted on a shaft. The turbine tends to give a greater shear than propellers. The shear can be increased by using diffuser rings that are perforated and surround the turbine so that the liquid from the turbine must pass through holes. The turbines can be used for both low-viscosity mixtures and medium-viscosity liquids. The degree of stirring and shear by propeller or turbine mixers depends upon several factors, such as the speed of rotation, pattern of liquid flow, position in the container, and baffles in the container.

Production sized mixers include high-powered propeller, shaft stirrers immersed in a tank, or self-contained units with propeller and paddle systems. The latter usually are constructed so that the contents of the tank either may be heated or cooled during the production process. Baffles often are built into a tank to increase mixing efficiency. Examples of two production dispersion mixers are shown in Figures 39-1 and 39-2.
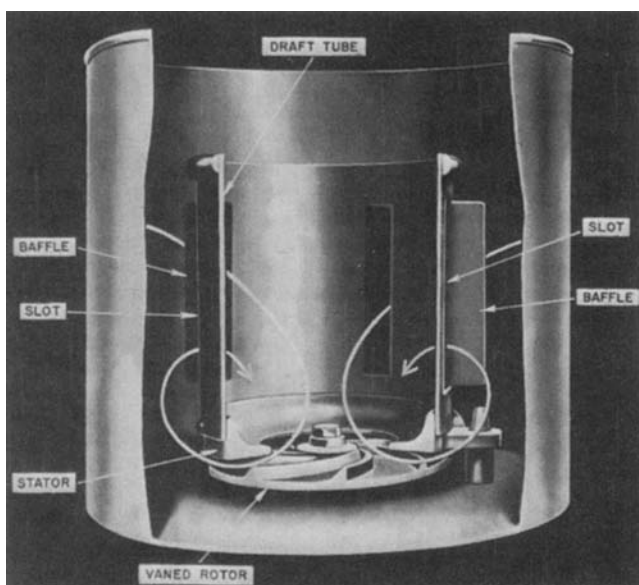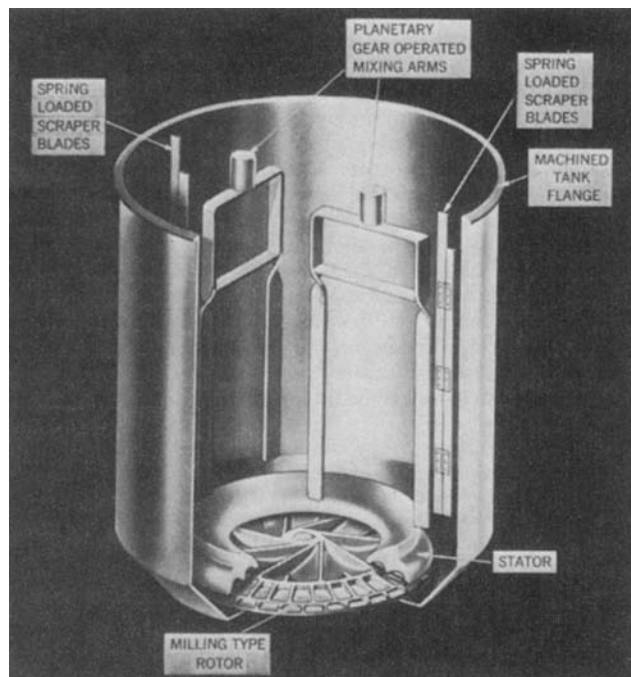

Figure 39-2. Standard paste-type dispersal mixer with cupped-rotor milling element and double-rotating mixing arm circulating element (courtesy, Abbe Eng).

Small electric mixers may be used to prepare emulsions at the prescription counter. They will save time and energy and produce satisfactory emulsions when the emulsifying agent is acacia or agar. The commercially available *Waring Blender* disperses efficiently by means of the shearing action of rapidly rotating blades. It transfers large amounts of energy and incorporates air into the emulsion. If an emulsion first is produced by using a blender of this type, the formulator must remember that the emulsion characteristics obtained in the laboratory will not necessarily be duplicated by the production-size equipment.

Colloid Mills

The principle of operation of the colloid mill is the passage of the mixed phases of an emulsion formula between a stator and a high-speed rotor revolving at speeds of 2,000 to 18,000 rpm. The clearance between the rotor and the stator is very small, but adjustable from 0.001 inches and up. The emulsion mixture, in passing between the rotor and stator, is subjected to a tremendous shearing action that effects a fine dispersion of uniform size. A colloid mill and various rotors are shown in Figures 39-3 and 39-4. The operating principle is the same for all, but each manufacturer incorporates specific features that result in changes in operating efficiency. The shearing forces applied in the colloid mill usually result in a temperature increase within the emulsion. It may be necessary, therefore, to use jacketed equipment to cool the emulsion during processing. Maa and Hsu have shown that droplet size of emulsions was mainly determined by shear force within the gap between the spinning rotor and stationary rotor.[81] Droplet size decreased with homogenization intensity and duration, increasing viscosity of the continuous phase, and with decreasing viscosity of the dispersed phase.

Colloid mills are used frequently for the comminution of solids and for the preparation of suspensions, especially suspensions containing solids that are not wetted by the dispersion medium, which are discussed later in this chapter.
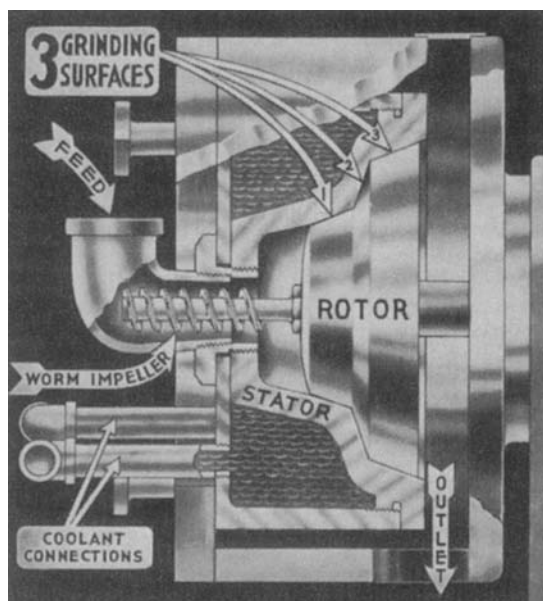

Figure 39-1. Standard slurry-type dispersal mixer with vaned-rotor mixing element and slotted draft-tube circulating element (courtesy, Abbe Eng).

Figure 39-3. A cross section of a colloid mill (courtesy, Tri-Homo).

Homogenizers

Impeller types of equipment frequently produce a satisfactory emulsion; however, for further reduction in particle size, homogenizers may be employed.[82] Homogenizers may be used in one of two ways:

1. The ingredients in the emulsion are mixed and then passed through the homogenizer to produce the final product.
2. A coarse emulsion is prepared in some other way and then passed through a homogenizer for the purpose of decreasing the particle size and obtaining a greater degree of uniformity and stability.

The mixed phases or the coarse emulsion are subjected to homogenization and are passed between a finely ground valve and seat under high pressure. This, in effect, produces an atomization that is enhanced by the impact received by the atomized mixture as it strikes the surrounding metal surfaces. They operate at pressures of 1,000 to 5,000 psi and produce some of the finest dispersions obtainable in an emulsion.

Figure 39-5 shows the flow through the homogenizing valve, the heart of the high pressure, APV Gaulin homogenizer. The product enters the valve seat at high pressure, flows through the region between the valve and the seat at high velocity with

Figure 39-4. Types of rotors used in colloid mills. These may be smooth (for most emulsions), serrated (for ointments and very viscous products), or of vitrified stone (for the paints and pigment dispersions) (courtesy, Tri-Homo).
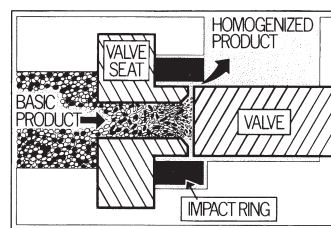

Figure 39-5. Material flow through a homogenizer (courtesy, APV Gaulin).

a rapid pressure drop, causing cavitation; subsequently, the mixture hits the impact ring causing further disruption and then is discharged as a homogenized product. It is postulated that circulation and turbulence are responsible mainly for the homogenization that takes place. Different valve assemblies, two-stage valve assemblies, and equipment with a wide range of capacities are available.

Two-stage homogenizers are constructed so that the emulsion, after treatment in the first valve system, is conducted directly to another where it receives a second treatment. A single homogenization may produce an emulsion that, although its particle size is small, has a tendency to clump or form clusters. Emulsions of this type exhibit increased creaming tendencies. This is corrected by passing the emulsion through the first stage of homogenization at a high pressure (eg, 3,000–5,000 psi) and then through the second stage at a greatly reduced pressure (eg, 1,000 psi). This breaks down any clusters formed in the first step.

For small-scale extemporaneous preparation of emulsions, the inexpensive hand-operated homogenizer is particularly useful. It is probably the most efficient emulsifying apparatus available to the prescription pharmacist. The two phases, previously mixed in a bottle, are hand-pumped through the apparatus. Recirculation of the emulsion through the apparatus will improve its quality.

A homogenizer does not incorporate air into the final product. Air may ruin an emulsion because the emulsifying agent is adsorbed preferentially at the air–water interface, followed by an irreversible precipitation termed *denaturization*. This is particularly prone to occur with protein emulsifying agents. Homogenization may spoil an emulsion if the concentration of the emulsifying agent in the formulation is less than that required to accommodate the increase in surface area produced by the process.

The temperature rise during homogenization is not very large. However, temperature does play an important role in the emulsification process. An increase in temperature will reduce the viscosity and, in certain instances, the interfacial tension between the oil and the water. There are, however, many instances, particularly in the manufacturing of cosmetic creams and ointments, where the ingredients will fail to emulsify properly if they are processed at too high a temperature. Emulsions of this type are processed first at an elevated temperature and then homogenized at a temperature not exceeding 40°C.

Homogenizers have been used most frequently with liquid emulsions, but now they may be used with suspensions, as the metal surfaces are formed from wear-resistant alloys that will resist the wear of solid particles contained in suspensions.

Ultrasonic Devices

The preparation of emulsions by the use of ultrasonic vibrations also is possible. An oscillator of high frequency (100–500 kHz) is connected to two electrodes between which is placed a piezoelectric quartz plate. The quartz plate and electrodes are immersed in an oil bath and, when the oscillator is operating, high-frequency waves flow through the fluid. Emulsification is accomplished by simply immersing a tube containing the

emulsion ingredients into this oil bath. Considerable research has been done on ultrasonic emulsification, particularly with regard to the mechanism of emulsion formation. The method has not been proven to be practical for large-scale production of emulsions, but evaluations are underway.[83]

Microfluidizers

Microfluidizers have been used to produce very fine particles. The process subjects the emulsion to an extremely high velocity through micro-channels in to an interaction chamber; as a result, particles are subjected to shear, turbulence, impact, and cavitation. Two advantages of this type of equipment are lack of contamination in the final product and ease of production scale up.

LIPOSOMES

Liposomes have been one of the most extensively studied drug delivery systems.[84–87] Liposomes, meaning lipid body, may be broadly described as small vesicles of a bilayer of phospholipid encapsulating an aqueous space ranging from about 0.03 to 10 μm in diameter. Generally, the lipid membrane of a liposome consists of a bilayer-forming amphiphile, cholesterol, and a charge-generating molecule. The lipid membrane encloses a discrete aqueous compartment. This structure presents an overall hydrophilic membrane-like assembly, in which the apolar or lipophilic portion of the amphiphilic molecule points inward while the polar or hydrophilic portion points outward of the lamellar structure. These characteristics make liposomes useful as drug delivery systems. The enclosed vesicles can encapsulate water soluble drugs in the aqueous spaces or lipid soluble drugs in the membranes. Liposomes have been administered parenterally, topically, and by inhalation.

Liposomes offer several advantages as drug delivery systems: (1) they are biologically inert and completely biodegradable; (2) they can be prepared in various sizes, charge, compositions, and surface morphology; (3) liposomes can encapsulate both water-soluble and water-insoluble drugs, including enzymes, hormones, and antibiotics; (4) encapsulated drugs are less susceptible to degradation; (5) organ-targeted drug delivery is possible since the entrapped drug is delivered intact to various tissues and cells after the liposome is destroyed; and (6) other tissues and cells of the body are protected from the drug until it is released by the liposomes, thus decreasing the drug's toxicity. The primary disadvantage of liposomes is their rapid removal from the blood following intravenous administration by cells of the reticuloendothelial system, particularly by Kupfer cells in the liver. Drug release is slowed by phagocytes through endocytosis, fusion, surface adsorption, or lipid exchange.

Several different amphiphiles have been investigated to create liposomal structures (vesicles). Only the bilayer-forming lipid is the essential part of the lamellar structure, and the other components are to impart specific characteristics. For example, cholesterol adds rigidity to the vesicular structure rendering it less permeable. Phospholipids such as phosphatidyl choline (lecithin) were the first amphiphiles used to produce bilayer structures to mimic cell membranes.

Liposomes can be prepared into several morphologies, which have been classified according to the vesicular shape. *Multilamellar vesicles* (MLV) were first prepared by Bangham and have multiple bilayer structures surrounding a relatively small internal core, much like an onion.[88] *Oligolamellar vesicles* (OLV) have large central aqueous cores surrounded by 2 to 10 bilayers. *Unilamellar vesicles* (ULV) have a single bilayer structure surrounding an internal aqueous core. Unilamellar vesicles can be prepared in a variety sizes: small unilamellar vesicles (20–40 nm), medium unilamellar vesicles (40–80 nm), large unilamellar vesicles (10–1,000 nm), and giant unilamellar vesicles (> 1,000 nm).

Drug release in the blood following intravenous administration ranges from a few minutes to several hours depending upon the nature and composition of the lipids, surface properties, and size. In general, smaller unilamellar vesicles show much longer half-lives than multilamellar vesicles and large unilamellar vesicles. Negatively charged liposomes are cleared more rapidly from the circulation than neutral or positively charged liposomes. Circulation can be prolonged by blocking the reticuloendothelial system and allowing the liposomes to interact with vascular endothelial cells and blood cells. These "stealth liposomes" were developed by coating the liposomes with polymers such as polyethylene glycol, enabling liposomes to evade detection by the body's immune system.

Preparation of Liposomes

Liposomes have been prepared using a number of techniques including solvent evaporation, sonication, supercritical fluid techniques, spray drying, extrusion, and homogenization. A combination of these methods is often used, and the drug is added during the formation process. In this method, the lipid is dissolved in an organic solvent such as acetone or chloroform. The solvent is evaporated leaving a thin, lipid film on the walls of the container. An aqueous solution of the drug is added and placed in an ultrasonic bath. The sound waves displace the lipid from the container walls, and they self-assemble into spheres or cylinders entrapping the aqueous drug solution inside. If the drug is lipophilic, it is incorporated into the lipid phase and will reside within the lipophilic bilayers. Several advances have been made in liposome preparation to better control stability and size.

Liposomal products are now commercially available. Amphotec (distributed by *InterMune*, manufactured by *Ben Venue Laboratories*) is Amphotericin B Cholesteryl Sulfate Complex for Injection. It is a sterile, pyrogen-free, lyophilized powder for reconstitution and intravenous (IV) administration. Amphotec consists of a 1:1 (molar ratio) complex of amphotericin B and cholesteryl sulfate. Upon reconstitution, Amphotec forms a colloidal dispersion of microscopic disc-shaped particles. Each 50 mg single dose vial contains amphotericin B, 50 mg; disodium edetate dihydrate, 0.372 mg; lactose monohydrate, 950 mg; and hydrochloric acid, qs. Amphotec is indicated for the treatment of invasive aspergillosis in patients where renal impairment or unacceptable toxicity precludes the use of amphotericin B deoxycholate in effective doses and in aspergillosis patients where prior amphotericin B deoxycholate therapy has failed. The drug is reconstituted with Sterile Water for Injection by rapidly adding the water to the vial; it is shaken gently by hand, rotating the vial until all the solids have dissolved. The fluid may be opalescent or clear.[89] For infusion, it is further diluted in 5% dextrose injection. The product should not be reconstituted with any fluid other than Sterile Water for Injection; do not reconstitute with dextrose or sodium chloride solutions. Also, for further dilution, it should not be admixed with sodium chloride or electrolytes. Solutions containing benzyl alcohol or any other bacteriostatic agent should not be used as they may cause precipitation. An inline filter should not be used, and the infusion admixture should not be mixed with other drugs. If infused using a y-injection site or similar device, flush the line with 5% dextrose injection before and after infusion of Amphotec. After reconstitution, the drug should be refrigerated and used within 24 hours; do not freeze. If further diluted with 5% dextrose injection, it should be refrigerated and used within 24 hours.

Doxil (*Ortho Biotech*) is doxorubicin hydrochloride encapsulated in stealth liposomes for intravenous administration. The product is provided as a sterile, translucent, red liposomal dispersion in a 10 mL glass, single use vial. Each vial contains 20 mg of doxorubicin HCl at a concentration of 2 mg/mL and a pH of 6.5. The stealth liposome carriers are composed of N-(carbonyl-methoxypolyethylene glycol 2000) - 1,2–distearoyl–

sn–glycerol–3 - phosphoethanolamine sodium salt (MPEG-DSPE), 3.19 mg/mL; fully hydrogenated soy phosphatidylcholine (HSPC), 9.58 mg/mL; and cholesterol, 3.19 mg/mL. Each mL also contains ammonium sulfate, approximately 2 mg; histidine as a buffer; hydrochloric acid and/or sodium hydroxide for pH control; and sucrose to maintain isotonicity. Greater than 90% of the drug is encapsulated in the Stealth liposomes. The stealth liposomes are specially formulated to circulate in the body "undetected" by the mononuclear phagocyte system for a prolonged circulation time of about 55 hours. This is accomplished by pegylation, or binding methoxypolyethylene glycol

on the surface of the liposomes. These liposomes are small, in the range of 100 nm in diameter. Doxil must be diluted in 250 mL of 5% dextrose injection prior to administration; once diluted it should be refrigerated and administered within 24 hours. It should not be mixed with any other diluent or any preservative-containing solution. It should not be used with inline filters. The product is not a clear solution but a red, translucent liposomal dispersion. Unopened vials should be stored in a refrigerator but freezing should be avoided, even though short-term freezing (less than 1 month) does not appear to adversely affect the product.

SUSPENSIONS

The physical chemist defines the word "suspension" as a two-phase system consisting of an undissolved or immiscible material dispersed in a vehicle (solid, liquid, or gas). A variety of dosage forms fall within the scope of this definition, but emphasis is placed on solids dispersed in liquids. In more specific terms, the pharmaceutical scientist differentiates between such preparations as suspensions, mixtures, magmas, gels, and lotions. In these preparations, the substance distributed is referred to as the dispersed phase and the vehicle is termed the dispersing phase or dispersion medium. In a general sense, each of these preparations represents a suspension, but the state of subdivision of the insoluble solid varies from particles that settle gradually on standing to particles that are colloidal in nature.

The particles of the dispersed phase vary widely in size, from large, visible particles to colloidal dimensions, which fall between 1.0 nm and 0.5 μm in size. Course dispersions contain particles usually 10–50 μm in size, and include suspensions and emulsions. Fine dispersions contain particles of smaller size, usually 0.5–10 μm. Magmas and gels represent such fine dispersions. Particles in a coarse dispersion have a greater tendency to separate from the dispersion medium than do the particles of a fine dispersion. Most solids in a dispersion tend to settle to the bottom of the container because their density is higher than the dispersion medium.

Suspensions have a number of applications in pharmacy. They are used to supply drugs to the patient in liquid form. Many people have difficulty swallowing solid dosage forms; consequently a liquid preparation has an advantage for these people. In addition, the dose of a liquid form may be adjusted easily to meet the patient's requirements. Thus, if the drug is insoluble or poorly soluble, a suspension may be the most suitable dosage form. If a drug is unstable in an aqueous medium, a different form of the drug, such as an ester or insoluble salt that does not dissolve in water, may be used in the preparation of a suspension. Drugs, such as antibiotics, that are unstable in the presence of an aqueous vehicle for extended periods of time are most frequently supplied as dry powder mixtures for reconstitution at the time of dispensing. This type of preparation is

designated in the USP by the title "for Oral Suspension." Suspensions that do not require reconstitution at the time of dispensing are simply designated as an "Oral Suspension." Examples of commercial products are presented in Table 39-7.

To improve the stability of an antibiotic such as ampicillin, formulations are made in such a way that the dispersion medium, water, is added upon dispensing to form a satisfactory suspension. Generally, the taste of pharmaceuticals can be improved if they are supplied in suspension form, rather than solutions; thus, chloramphenicol palmitate is used instead of the more soluble form, chloramphenicol. Another method to decrease the solubility of the drug is to replace part of the water with another appropriate liquid such as alcohol or glycerin. Insoluble drugs may be formulated as suspensions for topical use such as calamine lotion. Other preparations of suspensions, in addition to those noted above, include parenteral preparations (Chapter 41), ophthalmic preparations (Chapter 43), aerosol suspensions (Chapter 50), and medicated topicals (Chapter 44).

PHYSICAL CHARACTERISTICS OF SUSPENSIONS

Formulation of suspensions involves more than mixing a solid in a liquid. Knowledge of the behavior of particles in liquids, suspending agents, wetting agents, polymers, buffers, preservatives, flavors, and colors is required to produce an acceptable and satisfactory suspension. Suspensions should possess several basic chemical and physical properties. The dispersed phase should settle slowly, if at all, and be re-dispersed readily upon shaking. The solid particles should have a narrow particle size distribution, which does not cake on settling, and the viscosity should be such that the preparation pours easily. In addition, the product should have an elegant appearance, be resistant to microbial growth, and maintain its chemical stability.

Several factors influence the sedimentation rate of particles in a suspension. Stokes' law relates the diameter of the

Table 39-7. Examples of Commercial Suspensions

PRODUCT NAME	MANUFACTURER	ACTIVE INGREDIENT & DOSE	INDICATION
Carafate	Aventis	1 g sucralfate / 10 mL.	Antiulcer
Maalox	Novartis	225 mg aluminum hydroxide and 200 mg magnesium hydroxide / 5 mL	Antacid
Mepron	GlaxoSmithKline	750 mg atovaquone / 5 mL	Antiprotozoal
Mylanta Liquid	J&J-Merck	200 mg Aluminum Hydroxide, 200 mg Magnesium Hydroxide, and 20 mg simethicone / 5 mL	Antacid
Nystastin	Teva	100,000 units mycostatin / mL	Antifungal
Pepto-Bismol Liquid	Proctor & Gamble	262 mg bismuth subsalicyalte / 15 mL	Antidiarrheal
Pred-G ophthalmic suspension	Allegan	0.3% gentamicin and 1.0% prednisolone acetate	topical anti-inflammatory/ anti-infective
Viramune	Boehringer Ingelheim	50 mg of nevirapine / 5 mL	Antiviral

particles, the density of the particles and the medium, and the viscosity of the of the medium to the sedimentation rate:

$$\frac{dS}{dt} = \frac{d^2(\rho_P - \rho_M)g}{18\eta}$$

where
    dS/dt is the sedimentation rate,
    d is the diameter of the particles,
    ρ_P is the density of the particles,
    ρ_M is the density of the medium,
    g is the gravitational constant,
    η is the viscosity of the medium.

Stokes' equation was derived for an ideal situation with perfectly spherical particles in a very dilute suspension. It assumes the spherical particles settle without causing turbulence, without particle-to-particle collision, and without chemical or physical attraction or affinity for the dispersion medium. Obviously, the typical pharmaceutical suspension contains particles are irregularly shaped with a range of sizes, settling results in both turbulence and collision, and there is a reasonable affinity between the particles and suspension medium. However, the basic concepts of the equation offer an indication of the important variables for suspension of the particle and clues to formulation adjustments to decrease the rate of particle sedimentation.

Clearly, the sedimentation rate of large particles is greater than smaller particles, assuming all other factors remain constant. A slower rate of settling can be achieved by reducing particle size. Density also has a direct relationship with sedimentation rate: dense particles settle more rapidly than less dense particles. Most pharmaceutical suspensions are aqueous, and the density of the particles is generally greater than water; a desirable feature, since if they were less dense, they would float making a uniform product difficult to achieve.

The sedimentation rate is indirectly related to the medium viscosity, allowing the pharmaceutical scientist to manipulate settling by adjusting the viscosity of the medium. Settling is reduced by increasing the viscosity of the dispersion medium. One must keep in mind, a very high viscosity is not generally desirable, because it pours with difficulty and it is equally difficult to re-disperse. The viscosity characteristics of a suspension may be altered not only by the vehicle used, but also by the solids content. As the proportion of solid particles is increased in a suspension, so is the viscosity. In most cases, the physical stability of a pharmaceutical suspension is adjusted by the dispersed phase rather than through the dispersion medium. Generally, the dispersion medium supports the adjusted dispersed phase.

The most important consideration in formulation of suspensions is the size of the drug particles. In most pharmaceutical suspensions, the particle diameter is between 1 and 50 µm. The reduction in the particle size is beneficial to the stability of the suspension in that the rate of sedimentation is reduced as the particles are decreased in size. The reduction in particle size produces slow, more uniform rates of settling. However, reduction of the particle size to too great a degree of fineness should be avoided, since fine particles have a tendency to form a compact cake upon settling. The result may be that the cake resists breakup upon shaking and forms rigid aggregates of particles. Particle shape can also affect caking and product stability.[90]

Actions must be taken to prevent the agglomeration of particles into larger crystals or into masses, to avoid the formation of a cake. A common method to prevent rigid cohesion of small particles is through the intentional formation of a less rigid or loose aggregation of the particles by particle-to-particle bonding forces. An aggregation of this type is called a *floc* or *floccule*, in which particles form a lattice structure that resists complete settling and compaction. Flocs form a higher sediment volume than unflocculated particles, and the loose structure permits the aggregates to break up easily and redistribute with agitation. There are several methods of preparing flocculated suspensions, the choice depending on the drug and type of product desired. For example, clays such as bentonite are commonly used as flocculating agents in oral suspensions. The structure of bentonite and of other clays assists the suspension by helping to support the floc once formed. When clays are unsuitable, as in a parenteral suspension, a floc of the dispersed phase can be produced by an alteration in the pH of the preparation, generally to a region of low drug solubility. Electrolytes can also act as flocculating agents by reducing electrostatic interactions between the particles. Nonionic and ionic surfactants can also induce particle flocculation and increase the sedimentation volume.

Particle growth or *Ostwald ripening* is also a destabilizing process resulting from temperature fluctuations during storage. Temperature fluctuations may change particle size distribution and polymorphic form of a drug, if the solubility of the drug is temperature dependent. For example, if the temperature is raised, drug crystals may dissolve and form a supersaturated solution, which favor crystal growth on cooling. As the dissolved drug crystallizes out of solution, it will preferentially occur on the surface of a crystal in the suspension.

SUSPENSION INGREDIENTS

The external phase is usually water for oral preparations; however, other polar liquids such as glycerin or alcohol may be considered to control solubility, stability, and taste. The selection of the external phase is based upon taste, viscosity, density, and stability. Nonpolar liquids such as aliphatic hydrocarbons and fatty esters may be considered if the preparation is used for external purposes.

The main ingredients in a suspension are the drug and functional excipients that wet the drug, influence flocculation, control viscosity, adjust pH, and the external medium, usually water. In addition, flavoring, sweetening, and coloring agents and preservatives are employed. A *wetting agent* is a surfactant with an HLB value between 7 and 9. Surfactants with higher HLB values are recommended sometimes, such as polysorbates and poloxamers. They are employed at a low concentration (0.05–0.5%) to allow the displacement of air from hydrophobic material and permit the liquid, usually water, to surround the particles. If it is desirable to flocculate the particles, then flocculating agents are employed. Usually low concentrations, less than 1%, of electrolytes such as sodium or potassium chloride are employed to induce flocculation. Water-soluble salts possessing divalent or trivalent ions may be considered if the particles are highly charged.

Viscosity producing agents are generally polymers, including natural gums (acacia, xanthan) and cellulose derivatives, such as sodium carboxymethylcellulose and hydroxypropyl methylcellulose. These excipients are used at low concentrations to function as protective colloids, but at higher concentrations they function as viscosity increasing agents. At higher viscosity, the rate of settling of deflocculated particles is decreased providing additional stability to the flocculated suspension. The choice of an appropriate viscosity agent depends upon the use of the product (external or internal), processing equipment, and the duration of storage. Suspension preparations for internal use exhibiting good flow and suspending properties often contain sodium carboxymethylcellulose 2.5%, tragacanth 1.25%, or guar gum 0.5%. For external applications, Carbopol polymers have been successfully used. Other common viscosity-producing agents include acacia, methylcellulose, sodium alginate, or tragacanth.

Ideally, a suspension should be stable over a wide pH range. The chemical and / or physical stability of an active compound may occasionally require the pH of the medium to be maintained within a specified range. *Buffers* must be carefully considered so that they produce their intended effects without interference with other ingredients in the formulation. Buffers can influence the solubility of the active, preservative ionization and its activity, and ionic viscosity agents.

PREPARATION OF SUSPENSIONS

The preparation of suspensions involves several steps; the first is to obtain particles of the proper size, typically in the lower micrometer range. Oral preparations should not feel gritty, topical preparations should feel smooth to the touch, and injectables should not produce tissue irritation. Particle size and distribution also should be considered in terms of bioavailability, or from an in vitro perspective, the rate of release. Very small particles, less than 1 μm, will have a higher solubility than larger particles, but also have a faster *rate* of dissolution. Thus, particle size of the dispersed solid in a suspension can influence the rate of sedimentation, flocculation, solubility, dissolution rate, and ultimately, bioavailability.

Particle size reduction is generally accomplished by dry milling prior to the incorporation of the dispersed phase into the dispersion medium. *Milling* is the mechanical process of reducing particle size, which may be accomplished by a number of different types of machines. Hammer mills grind the powders by impact (Fig 39-6). Centrifugally rotating hammers or blades contact the particles and direct them against a screen, typically in the range of 4 to 325 mesh. The particles are forced through the screen, which regulates final particle size at the outlet of the milling chamber. The blade and screen act in conjunction to determine final product sizing, typically in the range of 10–50 μm.
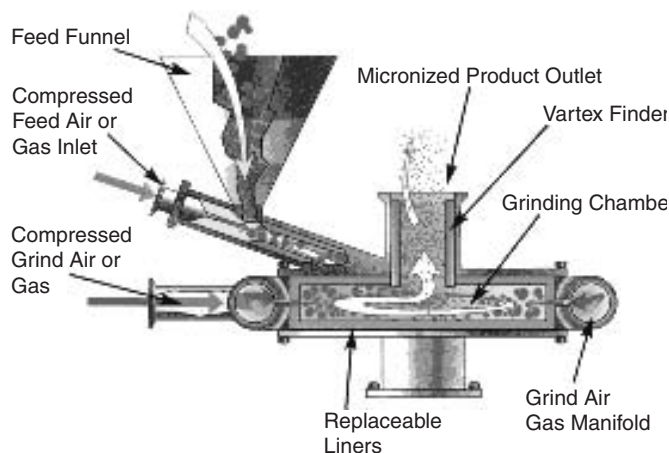
Fluid energy or jet mills produce particles under 25 μm through violent turbulence in high velocity air (Fig 39-7). Compressed air forms a high speed, jet stream which passes the feed funnel and draws powders into grinding chamber. Pulverizing nozzles are installed around the grinding chamber and inject additional high-speed air into the grinding chamber in a rotational direction. The centrifugal air-flow accelerates particles and reduces particle size by particle to particle impaction and friction. The air-flow drives large particles toward the perimeter, but small particles move toward the center where they exit through the outlet.

A ball mill contains a number of steel or ceramic balls in a rotating drum. The balls reduce the particle size to a 20 to 200 mesh by both attrition and impact. Roller mills have two or more rollers that revolve at different speeds, and the particles are reduced to a mesh of 20 to 200 by means of compression and a shearing action. See Chapter 37 (Powders) for a more detailed discussion on particle size reduction of solids.

In the pharmacy, ceramic mortar and pestle are better for grinding and reducing particle size than glass. After reducing particle size, the drug powder is wetted thoroughly with a small
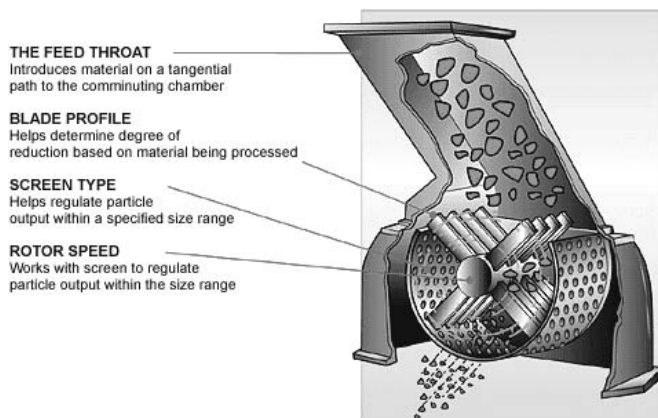
Figure 39-6. Material Flow through a Hammer Mill. Hammer mills operate by feeding material uniformly into a chamber in which a rotating blade assembly reduces the particles of the material by cutting or impacting them. The material discharges through a screen which regulates final particle size at the outlet of the milling chamber (courtesy, The Fitzpatrick Company).

THE FEED THROAT
Introduces material on a tangential path to the comminuting chamber

BLADE PROFILE
Helps determine degree of reduction based on material being processed

SCREEN TYPE
Helps regulate particle output within a specified size range

ROTOR SPEED
Works with screen to regulate particle output within the size range

Figure 39-7. Schematic of particle size reduction in a fluid energy or jet mill (courtesy, Sturtevant Inc.).

Feed Funnel

Micronized Product Outlet

Compressed Feed Air or Gas Inlet

Vartex Finder

Compressed Grind Air or Gas

Grinding Chamber

Grind Air Gas Manifold

Replaceable Liners

quantity of water miscible solvent, such as glycerin or alcohol, which reduces the interfacial tension. The suspending agent in the aqueous medium is then added. Alternately, the suspending agent can be triturated with the drug particles using a small quantity of glycerin or alcohol and then brought up to volume with the diluent water and triturated to a smooth uniform product.

On a large scale, the fine drug particles are treated with a small portion of water that contains the wetting agent and allowed to stand for several hours to release entrapped air. At the same time, the suspending agent should be dissolved or dispersed in the main portion of the external phase and allowed to stand until complete hydration takes place. Subsequently, wetted drug particles should be added slowly to the main portion of the dissolved suspending agent. Other excipients such as electrolytes or buffers should be carefully introduced. The preservatives, flavoring agents, and coloring agents are added last. Finally, the formulation is processed with homogenizers, ultrasonic devices, or colloid mills to produce a uniform product.

A procedure for the preparation of Trisulfapyrimidines Oral Suspension is given below.

Trisulfapyrimidines Oral Suspension

Veegum	1.00 g
Syrup USP	90.60 g
Sodium Citrate	0.78 g
Sulfadiazine	2.54 g
Sulfamerazine	2.54 g
Sulfamethazine	2.54 g

Add the Veegum slowly and with continuous stirring to the syrup. Incorporate the sodium citrate into the Veegum–syrup mixture. Premix the sulfa drugs, add to the syrup, stir, and homogenize. Add sufficient 5% citric acid to adjust the pH of the product to 5.6. A preservative and a flavoring agent may be added to the product.

QUALITY CONSIDERATIONS

The quality of the suspension can be determined in a number of ways. Particle size, particle size distribution, and particle shape are often determined using photo microscopy or laser light diffraction techniques. Physical stability, the degree of settling, or flocculation may be determined using a device to measure the zeta potential. Viscosity may be determined by instruments such as the Brookfield viscometer or of the cone and plate configuration. Microbiological as well as stability testing according to ICH guidelines should be performed to determine the efficiency of the preservative and the appropriateness of the formulation with respect to time, temperature, and relative humidity.[91]

EXTEMPORANEOUS PREPARATIONS FROM TABLETS AND CAPSULES

Occasionally, it is necessary to prepare a liquid formulation of a drug to meet certain patient requirements. Consequently, patients who are unable to swallow solid medications, require a different route of administration or different dosing strength present a special need. Thus, the pharmacist may have to extemporaneously compound a liquid product. If the pure drug is available, it should be used to prepare the liquid dosage form. If it is necessary to prepare a liquid dosage form from tablets or capsules, a suspension is formed if either the drug or one of the excipients in the tablets or capsules is insoluble. Insoluble excipients in these dosage forms include disintegrants, lubricants, glidants, colors, diluents, and coatings. Consequently, although the drug may be soluble in water, many excipients are not. It is preferable to use the contents of capsules, or tablets that are not coated. If coated, tablets with a water-soluble coat are preferred to those with functional enteric coatings and the like. In any case, the contents of the capsules or the tablets should be ground finely with a ceramic mortar and pestle and then wetted using alcohol or glycerin.

Preservatives may be included in the liquid formulation to enhance the stability. However, preservatives have been found to cause serious adverse effects in infants. Benzyl alcohol should be omitted from neonatal formulations because it can cause a gasping syndrome characterized by a deterioration of multiple organ systems and eventually death. Propylene glycol has also been implicated to cause seizures and stupor in some preterm infants. Thus, formulations for neonates should be purposely kept simple, and not compounded to supply more than just a few days of medicine

Finally, it may be desirable to use a hand homogenizer to prepare a more suitable product. Some drugs that have been formulated in this manner include clonidine hydrochloride and simple syrup,[92] cefuroxime axetil in an orange syrup vehicle,[93] and famotidine in cherry syrup.[94] Many other examples may be found in current hospital and community pharmacy journals such as the *American Journal of Hospital Pharmacy, Canadian Journal of Hospital Pharmacy*, *U.S. Pharmacist*, *International Journal of Pharmaceutical Compounding,* and *Drug Development and Industrial Pharmacy*. Frequently, stability data and, occasionally, bioavailability and/or taste data are provided.

To minimize stability problems of the extemporaneously prepared product, it should be placed in air-tight, light-resistant containers and stored in the refrigerator by the patient. Because it is a suspension, the patient should be counseled to shake it well prior to use and to be aware of any change that might indicate a stability problem with the formulation.

Tortorici reports an example of an extemporaneous suspension of cimetidine tablets that retained its potency at 40° over 14 days.[95] Twenty-four, 300 mg cimetidine tablets are compounded with 10 mL of glycerin and 120 mL of simple syrup. The tablets are triturated to a fine powder using a mortar, the mixture is levigated with the glycerin, and the simple syrup added. The suspension is mixed well, placed in a blender until smooth, and then refrigerated.

SUSTAINED RELEASE SUSPENSIONS

Sustained release suspensions represent a very specialized class of preparation. Sustained release, oral suspensions with morphine,[96] nonsteroidal anti-inflammatory agents,[97] and other drugs[98] have been described in the literature. However, limited commercial success has been achieved due to the difficulty in maintaining the stability. Formulation research for sustained release suspensions has focused on the similar technologies used in preparing sustained release tablets and capsules. *Celltech* licenses the Tussionex Pennkinetic system, which uses a combination of ion exchange resin and particle coating.[99] This novel system exploits the likelihood of complexation between ionic drugs and ion-exchange resins, which are then coated with ethyl cellulose. When administered orally, the coated particles with encapsulated drug adsorbed onto the resin are slowly released by an ion exchange process.

Durect markets the SABER system for sustained release suspension applications. SABER uses a non-polymeric, non-water-soluble high-viscosity liquid carrier material (>5,000 cPs. at 37°C), such as sucrose acetate isobutyrate (SAIB), to provide controlled release of active ingredients.[100] The drug is mixed with a small amount of a pharmaceutically acceptable solvent to form a low viscosity solution or suspension, which is then mixed with the high viscosity carrier. The resulting suspension can be administered via injection, orally, or as an aerosol, forming an adhesive, biodegradable depot upon contact with tissues. After administration of the SABER formulation, the solvent diffuses away, leaving a viscous, adhesive matrix of the three components—SAIB, drug, and any additives. The release rate can be easily modified by the ratio of non-polymeric, non-water-soluble high-viscosity liquid carrier material present in the formulation. Extended systemic and local delivery for durations of 1 day to 3 months from a single injection has been demonstrated.

GELS AND MAGMAS

Gels are defined by the USP as:

> "...semisolid systems consisting of either suspensions made up of small inorganic particles or large organic molecules interpenetrated by a liquid. Where the gel mass consists of a network of small discrete particles, the gel is classified as a two-phase system. In a two-phase system, if the particle size of the dispersed phase is relatively large, the gel mass is sometimes referred to as a magma. Both gels and magmas may be thixotropic, forming semisolids on standing and becoming liquid on agitation.
>
> Single-phase gels consist of organic macromolecules uniformly distributed throughout a liquid in such a manner that no apparent boundaries exist between the dispersed macromolecules and the liquid. Single-phase gels may be made from synthetic macromolecules or from natural gums. The latter preparations are also called mucilages. Although these gels are commonly aqueous, alcohols and oils may be used as the continuous phase. For example, mineral oil can be combined with a polyethylene resin to form an oleaginous ointment base.
>
> Gels can be used to administer drugs topically or into body cavities."

Gels are also defined as semi-rigid systems in which the movement of the dispersing medium is restricted by an interlacing three-dimensional network of particles or solvated macromolecules in the dispersed phase. Physical and / or chemical cross-linking may be involved. The interlacing and consequential internal friction is responsible for increased viscosity and the semisolid state.

Some gel systems are clear and others are turbid, since the ingredients involved may not be completely soluble or insoluble, or they may form aggregates, which disperse light. The concentration of the gelling agents is generally less than 10%, and usually in 0.5 to 2.0% range. Gels in which the macromolecules are distributed throughout the liquid in such a manner that no apparent boundaries exist between them and the liquid are called single-phase gels. In instances in which the gel mass consists of floccules of small distinct particles, the gel is classified as a two-phase system and frequently called a magma or a milk. Gels and magmas are considered colloidal dispersions since they each contain particles of colloidal dimension

Different types of colloidal dispersions have been given specific names. For instance, *sol* is a general term designating a dispersion of a solid substance in a liquid, a solid, or a gaseous

dispersion medium. However, more often than not it is used to describe the solid liquid dispersion system. A prefix such as hydro- for water (*hydrosol*) or alco- for alcohol (*alcosol*) is used to specify the medium. Similarly, *aerosol* has similarly been developed to indicate a dispersion of a solid or a liquid in a gaseous phase.

The generally accepted size range for a substance "colloidal" is when particles fall between 1 nm and 0.5 µm. One difference between colloidal dispersions and true solutions is the larger particle size of the dispersed phase in colloidal systems. The optical properties of the two systems are also different. True solutions do not scatter light and therefore appear clear, but colloidal dispersions contain discrete particles scatter light.

Gelling Agents

Several compendial materials function as gelling agents, including acacia, alginic acid, bentonite, carbomer, carboxymethylcellulose sodium, cetostearyl alcohol, colloidal silicon dioxide, ethylcellulose, gelatin, guar gum, hydroxyethylcellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, magnesium aluminum silicate, maltodextrin, methylcellulose, polyvinyl alcohol, povidone, propylene carbonate, propylene glycol alginate, sodium alginate, sodium starch glycolate, starch, tragacanth, and xanthan gum.

Alginic acid is refined from seaweed. It is a tasteless, practically odorless, white to off-white colored, fibrous powder. It is used in concentrations between 1% and 5% as a thickening agent, and swells in water to about 200 times its own weight without dissolving. Alginic acid can be cross-linked by addition of calcium salts, resulting in substantially higher viscosity. Sodium alginate produces a gel at concentrations up to 10%. Aqueous preparations are most stable between pH values of 4–10; below pH 3, alginic acid is precipitated. Sodium alginate gels for external use should be preserved.

Carbomer resins are high molecular weight, acrylic acid-based polymers. The pH of 0.5% and 1.0% aqueous dispersions are 2.7–3.5 and 2.5–3.0, respectively. There are many carbomer resins, with viscosity ranges available from 0 to 80,000 cPs., depending upon the pH to which it is neutralized. In addition to thickening, suspending, and emulsifying in both oral and topical formulations, carbomers are also used to provide sustained release properties in both the stomach and intestinal tract for commercial products. Alcohol is often added to carbomer gels to decrease their viscosity. Carbomer gel viscosity is also dependent upon the presence of electrolytes and the pH. Generally, a rubbery mass forms if greater than 3% electrolytes are added. Carbomer preparations are primarily used in aqueous systems, although other liquids can be used. In water, a single particle of carbomer will wet very rapidly but, like many other powders, carbomer polymers tend to form clumps of particles when haphazardly dispersed in polar solvents. Rapid dispersion of carbomers can be achieved by adding the powder very slowly into the vortex of the liquid that is very rapidly stirred. A neutralizer is added to thicken the gel after the carbomer is dispersed. Sodium hydroxide or potassium hydroxide can be used in carbomer dispersions containing less than 20% alcohol. Triethanolamine will neutralize carbomer resins containing up to 50% ethanol.

Carboxymethylcellulose (CMC) produces gels when used in concentrations of 4% to 6% of the medium viscosity grade. Glycerin may be added to prevent drying. Precipitation will occur at pH values less than 2, it is most stable at pH levels between 2 and 10, with maximum stability at pH 7 to 9. It is incompatible with ethanol. Sodium carboxymethylcellulose (NaCMC) is soluble in water and should be dispersed with high shear in cold water before the particles hydrate and swell. Once the powder is well dispersed, the solution is heated with moderate shear to about 60°C for fastest dissolution. These colloidal dispersions are sensitive to pH and the viscosity of the product decreases below pH 5 or above pH 10.

Tragacanth gum has been used to prepare gels that are stable at a pH range of 4–8. These gels must be preserved or sterilized by autoclaving. Tragacanth often lumps when added to water, thus, aqueous dispersions are prepared by adding the powder to rapidly mixed water. Also, lumps are also prevented by wetting the gum with ethanol, glycerin, or propylene glycol.

Colloidal silicon dioxide can be used to prepare transparent gels when used with other ingredients of similar refractive index. Colloidal silicon dioxide adsorbs large quantities of water without liquefying, and its viscosity is largely independent of temperature. Changes in pH affect the viscosity: it is most effective at pH values up to about 7.5. Colloidal silicon dioxide (fumed silica) will form a hydrophobic gel when combined with 1-dodecanol and n-dodecane. These are prepared by adding the silica to the vehicle and sonicating for about 1 minute to obtain a uniform dispersion, sealing, and storing at about 40°C overnight.

Gelatin gels are prepared by dispersing gelatin in hot water followed by cooling. Alternatively, gelatin can be wetted with an organic liquid such as ethyl alcohol or propylene glycol followed by the addition of the hot water and cooling. Magnesium aluminum silicate forms thixtropic gels at concentrations of about 10%. The material is inert and has few incompatibilities but is best used above pH 3.5. It may bind to some drugs and limit their availability.
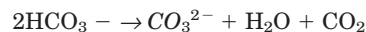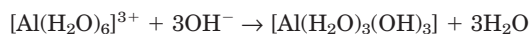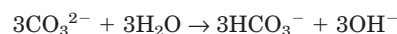
Methylcellulose forms gels at concentrations up to about 5%. Since methylcellulose hydrates slowly in hot water, the powder is dispersed with high shear at 80–90°C in a portion of water. Once the powder is finely dispersed, the remaining water is added with moderate stirring. Alcohol or propylene glycol is often used to help wet the powders. High electrolyte concentrations will salt out the polymer, ultimately precipitating the polymer.

Poloxamer gels are made from selected forms of polyoxyethylene-polyoxypropylene copolymers in concentrations ranging from 15% to 50%. Poloxamers are white, waxy, free-flowing granules that are practically odorless and tasteless. Aqueous solutions of poloxamers are stable in the presence of acids, alkalis, and metal ions. Polyvinyl alcohol (PVA) is used at concentrations of about 2.5% in the preparation of various jellies, which dry rapidly when applied to the skin. Borax is a often used to gel PVA solutions. For best results, disperse PVA in cold water, followed by hot water. It is less soluble in the cold water.

Povidone, in the higher molecular weight forms, can be used to prepare gels in concentrations up to about 10%. It has the advantage of being compatible in solution with a wide range of inorganic salts, natural and synthetic resins, and other chemicals. It has also been used to increase the solubility of a number of poorly soluble drugs.
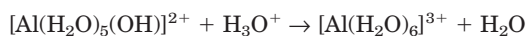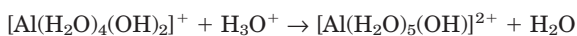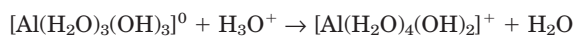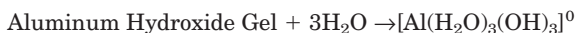
Two-Phase Gels

Two-phase gels containing bentonite may be used as a base for topical preparations such as plaster and ointment. Aluminum Hydroxide Gel, USP is an example of a two-phase gel. The USP states that "Aluminum Hydroxide Gel is a suspension of amorphous aluminum hydroxide in which there is a partial substitution of carbonate for hydroxide." The gel is usually prepared by the interaction of a soluble aluminum salt, such as a chloride or sulfate, with ammonia solution, sodium carbonate, or bicarbonate. The reactions that occur during the preparation are

$$3CO_3^{2-} + 3H_2O \rightarrow 3HCO_3^- + 3OH^-$$

$$[Al(H_2O)_6]^{3+} + 3OH^- \rightarrow [Al(H_2O)_3(OH)_3] + 3H_2O$$

$$2HCO_3^- \rightarrow CO_3^{2-} + H_2O + CO_2$$

The physical and chemical properties of the gel will be affected by the order of addition of reactants, pH of precipitation, temperature of precipitation, concentration of the reactants, the reactants used, and the conditions of aging of the precipitated gel.

Aluminum Hydroxide Gel is soluble in acidic (or very strongly basic) media. The mechanism in acidic media is

$$\text{Aluminum Hydroxide Gel} + 3H_2O \rightarrow [Al(H_2O)_3(OH)_3]^0$$

$$[Al(H_2O)_3(OH)_3]^0 + H_3O^+ \rightarrow [Al(H_2O)_4(OH)_2]^+ + H_2O$$

$$[Al(H_2O)_4(OH)_2]^+ + H_3O^+ \rightarrow [Al(H_2O)_5(OH)]^{2+} + H_2O$$

$$[Al(H_2O)_5(OH)]^{2+} + H_3O^+ \rightarrow [Al(H_2O)_6]^{3+} + H_2O$$

It is unlikely that the last reaction given proceeds to completion. Because the activity of the gel is controlled by its insolubility. Further, because a certain quantity of insoluble gel always is available, the neutralizing capability of the gel extends over a considerable period of time.

Aluminum hydroxide gels also may contain peppermint oil, glycerin, sorbitol, sucrose, saccharin, and various preservatives. Sorbitol improves the acid-consuming capacity by inhibiting a secondary polymerization that takes place on aging. In addition, polyols such as mannitol, sorbitol, and inositol have been shown to improve the stability of aluminum hydroxide and aluminum hydroxycarbonate gels.[101]

Single-Phase Gels

Single-phase gels are used more frequently in pharmacy for several reasons: semisolid state, high degree of clarity, ease of application, and ease of removal and use. The gels often provide a faster release of drug substance, independent of the water solubility of the drug, as compared to creams and ointments.

Some recent gel formulations include ophthalmic preparations of pilocarpine, carbachol, and betamethasone valerate; topical preparations for burn therapy, anti-inflammatory treatment, musculoskeletal disorders, and acne; peptic ulcer treatment with sucralfate gel; and bronchoscopy using lidocaine. Gels may be used as lubricants for catheters and bases for patch testing, and sodium chloride gels are used for electrocardiography.

Some gel formulation examples are provided below.

Methylcellulose and Carbomer Gel Base

Methylcellulose, 4000 cps	1.0 %
Carbomer 934	0.35 %
1 N Sodium hydroxide solution	qs to pH 7
Propylene glycol	16.7 %
Methyl paraben	0.015%
Purified water,	qs 100%

Disperse the methylcellulose in a portion of hot (80–90°C) water. Cool to room temperature, and disperse the Carbomer 934 in the gel using a bladed impeller. Adjust the pH of the dispersion to 7.0 by adding sufficient 1 N sodium hydroxide solution. Dissolve the methylparaben in the propylene glycol. Mix the methylcellulose, Carbopol 934 and propylene glycol fractions using caution to avoid incorporating air.

Sodium Alginate Gel Base

Sodium Alginate	10 g
Glycerin	10 g
Methyl Hydroxybenzoate	0.2 g
A soluble calcium salt (calcium gluconate)	0.5 g
Purified Water,	to make 100 mL

Place a portion of water in a beaker and add the glycerin and preservative. Stir this solution with a high speed mixer and add the sodium alginate. The calcium salt is added next, which increases the viscosity. Continue mixing until the preparation is homogeneous. The preparation should be stored in a tightly sealed wide mouth jar or tube.

Carbomer Gel

Carbomer 934	2 g
Triethanolamine	1.65 mL
Methyl Paraben	0.2 g
Propyl Paraben	0.05 g
Purified Water,	to make 100 mL

The parabens are dissolved in 95 mL of water with the aid of heat and allowed to cool. Carbomer 934 is added in small amounts to the solution using a high speed mixer until a smooth dispersion is obtained. The preparation is allowed to stand, permitting entrapped air to separate. Then the neutralizing agent, triethanolamine, is added very slowly to avoid entrapping air. Finally, the remaining water is then incorporated.

LOTIONS

Lotions are not defined specifically in the USP, but a broad definition describes them as either liquid or semi-liquid preparations that contain one or more active ingredients in an appropriate vehicle. Lotions may contain antimicrobial preservatives and other appropriate excipients such as stabilizers. Lotions are intended to be applied to the unbroken skin without friction. Lotions are usually suspensions of solids in an aqueous medium. Some lotions are, in fact, emulsions or solutions.

Even though lotions usually are applied without friction, the insoluble matter should be divided very finely. Particles approaching colloidal dimensions are more soothing to inflamed areas and effective in contact with infected surfaces. A wide variety of ingredients may be added to the preparation to produce better dispersions or to accentuate its cooling, soothing, drying, or protective properties. Bentonite is a good example of a suspending agent used in the preparation of lotions. Methylcellulose or sodium carboxymethylcellulose, for example, will localize and hold the active ingredient in contact with the affected site and at the same time be rinsed off easily with water. A formulation containing glycerin will keep the skin moist for a considerable period of time. The drying and cooling effect of a lotion may be accentuated by adding alcohol to the formula.

Dermatologists frequently prescribe lotions containing anesthetics, antipruritics, antiseptics, astringents, germicides, protectives, or screening agents, to be used in treating or preventing various types of skin diseases and dermatitis. Antihistamines, benzocaine, calamine, resorcin, steroids, sulfur, zinc oxide, betamethasone derivatives, salicylic acid, safflower oil, minoxidil, and zirconium oxide are ingredients common in lotions.

Lotions may be prepared by triturating the ingredients to a smooth paste and then adding the remaining liquid phase with trituration. High-speed mixers or colloid mills produce better dispersions and, therefore, are used in the preparation of larger quantities of lotion. Calamine Lotion USP is the classic example of this type of preparation and consists of finely powdered, insoluble solids held in more or less permanent suspension by the presence of suspending agents and/or surface-active agents. The formula and the method of preparation of Calamine Lotion, USP follows.

Calamine Lotion, USP

Calamine	80 g
Zinc Oxide	80 g
Glycerin	20 mL
Bentonite Magma	250 mL
Calcium Hydroxide Topical Solution	qs 1,000 mL

Dilute the bentonite magma with an equal volume of calcium hydroxide topical solution. Mix the powder intimately with the glycerin and about 100 mL of the diluted magma, triturating until a smooth, uniform paste is formed. Gradually incorporate the remainder of the diluted magma. Finally add enough calcium hydroxide topical solution to make 1000 mL, and shake well. If a more viscous consistency in the Lotion is desired, the quantity of bentonite magma may be increased to not more than 400 mL.

Many investigators have studied Calamine Lotion, and this has led to the publication of many formulations, each possessing certain advantages over the others, but none satisfying the col-

lective needs of all dermatologists. Formulations containing hydrated microcrystalline cellulose and carboxymethylcellulose have a slower rate of sedimentation than the official preparation.

Although most lotions are prepared by trituration, some lotions are formed by chemical interaction in the liquid. White Lotion, USP is an example.

White Lotion

Zinc Sulfate	40 g
Sulfurated Potash	40 g
Purified Water,	qs 1,000 mL

Dissolve the zinc sulfate and the sulfurated potash separately, each in 450 mL of purified water, and filter each solution. Add slowly the sulfurated potash solution to the zinc sulfate solution with constant stirring. Then add the required amount of purified water, and mix.

Benzyl Benzoate Lotion USP is an example of a lotion that is also an emulsion. The formula and method of preparation are:

Benzyl Benzoate	250 mL
Triethanolamine	5 g
Oleic Acid	20 g
Purified Water	qs 1,000 mL

Mix the triethanolamine with the oleic acid, add the benzyl benzoate, and mix. Transfer the mixture to a suitable container of about 2000 mL capacity, add 250 mL of purified water, and shake the mixture thoroughly. Finally add the remaining purified water, and again shake thoroughly.

Triethanolamine forms a soap with the oleic acid and functions as the emulsifying agent to form a stable product. This type of emulsifying agent is almost neutral in water and gives a pH of about 8 and thus should not irritate the skin.

Certain lotions tend to separate or stratify on long standing, and they require a label directing that they be shaken well before each use. All lotions should be labeled "For External Use Only." Microorganisms may grow in certain lotions if no preservative is included. Care should be taken to avoid contaminating the lotion during preparation, even if a preservative is present.

Milk of Magnesia USP is a suspension of magnesium hydroxide containing approximately 80 mg of $Mg(OH)_2$ per milliliter. The specifications for double strength or triple strength are that these products should contain approximately 160 mg or 240 mg of $Mg(OH)_2$ per mL, respectively. It has an unpleasant, alkaline taste that can be masked with 0.1% citric acid (to reduce alkalinity) and 0.05% of a volatile oil or a blend of volatile oils. Magnesium hydroxide is prepared by the hydration of magnesium oxide.

For the most part, magmas are intended for internal use, although Bentonite Magma is used primarily as a suspending agent for insoluble substances for local application and occasionally for internal use. All magmas require a "Shake Well" label and "Avoid Freezing."

EXTRACTS

Extraction, as the term is used pharmaceutically, involves the separation of medicinally active portions of plant or animal tissues from the inactive or inert components by using selective solvents in standard extraction procedures. The products obtained from plants are relatively impure liquids, semisolids, or powders intended only for oral or external use. These include classes of preparations known as decoctions, infusions, fluidextracts, tinctures, pilular (semisolid) extracts, and powdered extracts. Such preparations popularly have been called galenicals, after Galen, the 2nd century Greek physician.

Extraction continues to be of considerable interest in order to obtain improved yields of drugs derived from plant and animal sources. For example, extraction of digitalis glycosides has been carried out using super critical carbon dioxide.[102] Other techniques include ultrasonics, rotary-film evaporators, hydrodistillation, liquid chromatography, multiple-solvent extraction, countercurrent extraction, and gravitation dynamics.

This discussion is concerned primarily with basic extraction procedures for crude drugs to obtain the therapeutically desirable portion and eliminate the inert material by treatment with a selective solvent, known as the menstruum. Extraction differs from solution in that the presence of insoluble matter is implied in the former process. The principal methods of extraction are maceration, percolation, digestion, infusion, and decoction. The quality of the finished product can be enhanced by standardizing primary extracts and carrying out analytical assays during production on the raw materials, intermediate products, and manufacturing procedures.

The processes of particular importance, insofar as the USP is concerned, are those of maceration and percolation, as described specifically for Belladonna Extract USP and Cascara Sagrada Extract USP. Most pharmacopeias refer to such processes for extraction of active principles from crude drugs. The USP provides general directions for both maceration and percolation under the heading of *Tinctures*.

Techniques of extraction continue to be investigated and applied to obtain higher yields of the active substance from natural sources. Some of these methods include the use of different grinding and shearing processes of plants, use of specific membranes for extraction, and different extraction procedures such as distillation, digestion, percolation, and microwaves.

MACERATION—In this process the solid ingredients are placed in a stoppered container with 750 mL of the prescribed solvent and allowed to stand for a period of at least 3 days in a warm place with frequent agitation, until soluble matter is dissolved. The mixture is filtered and, after most of the liquid has drained, the residue on the filter is washed with sufficient quantity of the prescribed solvent or solvent mixture; the filtrates are combined to produce 1000 mL.

PERCOLATION—The ground solids are mixed with the appropriate quantity of the prescribed solvent to make it evenly and uniformly damp. It is allowed to stand for 15 min, then transferred to a percolator and packed. Sufficient prescribed solvent is added to saturate the solids. The top is placed on the percolator, and when the liquid is about to drip from the apparatus, the lower opening is closed. The solids are allowed to macerate for 24 hours or for the specified time. If no assay is directed, the percolation is allowed to proceed slowly or at the specified rate gradually adding sufficient solvent to produce 1000 mL of solution. If an assay is required, only 950 mL of percolate are collected and mixed and a portion assayed as directed. The rest of the percolate is diluted with the solvent to produce a solution that conforms to the required standard and then mixed.

DIGESTION—This is a form of maceration in which gentle heat is used during the process of extraction. It is used when moderately elevated temperature is not objectionable and the solvent efficiency of the menstruum is increased thereby.

INFUSION—An infusion is a dilute solution of the readily soluble constituents of crude drugs. Fresh infusions are prepared by macerating the solids for a short period of time with either cold or boiling water. The USP has not included infusions for some time.

DECOCTION—This once popular process extracts water-soluble and heat stable constituents from crude drugs by boiling in water for 15 min, cooling, straining, and passing sufficient cold water through the drug to produce the required volume.

EXTRACTIVE PREPARATIONS

After a solution of the active constituents of a crude drug is obtained by maceration or percolation, it may be ready for use as a medicinal agent, as with certain tinctures or fluidextracts, or it may be processed further to produce a solid or semisolid extract.

Tinctures

Tinctures are defined in the USP as being alcoholic or hydroalcoholic solutions prepared from vegetable materials or from chemical substances, an example of the latter being Iodine Tincture. Traditionally, tinctures of potent vegetable drugs essentially represent the activity of 10 g of the drug in each 100 mL of tincture, the potency being adjusted following assay. Most other tinctures of vegetable drugs represent the extractive from 20 g of the drug in 100 mL of tincture.

The USP specifically describes two general processes for preparing tinctures, one by percolation and the other by maceration. Percolation includes a modification so that tinctures that require assay for adjustment to specified potency thus may be tested before dilution to final volume. Belladonna Tincture, USP is prepared in this manner. Compound Benzoin Tincture USP and Sweet Orange Peel Tincture, USP are prepared by the maceration procedure.

Fluidextracts

The USP defines fluidextracts as being liquid preparations of vegetable drugs, containing alcohol as a solvent or as a preservative, or both, so made that, unless otherwise specified in an individual monograph, each milliliter contains the therapeutic constituents of 1 g of the standard drug that it represents.

Extracts

Extracts are defined in the USP as concentrated preparations of vegetable or animal drugs obtained by removal of the active constituents of the respective drugs with suitable menstrua, evaporation of all or nearly all of the solvent, and adjustment of the residual masses or powders to the prescribed standards. There are three forms of extracts: semiliquids or liquids of syrupy consistency, plastic masses (known as *pilular* or *solid extracts*), and dry powders (known as *powdered extracts*). Extracts, as concentrated forms of the drugs from which they are prepared, are used in a variety of solid or semisolid dosage forms. The USP states that pilular extracts and powdered extracts of any one drug are interchangeable medicinally, but each has its own pharmaceutical advantages. Pilular extracts, so-called because they are of a consistency to be used in pill masses and made into pills, are also suited for use in ointments and suppositories. Powdered extracts are better suited for incorporation into a dry formulation, as in capsules, powders, or tablets. Semiliquid extracts, or extracts of a syrupy consistency, may be used in the manufacture of some pharmaceutical preparations.

Most extracts are prepared by extracting the drug by percolation. The percolate is concentrated, generally by distillation under reduced pressure. The use of heat is avoided where possible because of potential injurious effect on active constituents. Powdered extracts that are made from drugs that contain inactive oily or fatty matter may have to be defatted or prepared from defatted drug. Pure Glycyrrhiza Extract USP is an example of a pilular extract and Belladonna Extract USP is an example of a powdered extract.

BIBLIOGRAPHY

General

Nielloud F, Marti-Mestres G, eds. *Pharmaceutical emulsions and suspensions*. New York: Marcel Dekker, 2000.
Lieberman HA, Rieger MM, Banker GS, eds. *Pharmaceutical Dosage Forms. Volume 3: Disperse Systems,* 2nd ed. New York: Marcel Dekker, 1998.
Lachman L, Liebermann HA, Kanig J, eds. *The Theory and Practice of Industrial Pharmacy,* 3rd ed. Philadelphia: Lea & Febiger, 1986.

Solutions, Emulsions and Suspensions

Becher P. *Emulsions: Theory & Practice*, 3rd ed. New York: Oxford, 2001.
Becher P. *Encyclopedia of Emulsion Technology*. New York: Marcel Dekker, 1983.
Kreuter J. *Colloidal Drug Delivery Systems*. New York: Marcel Dekker, 1994.
Osborne DW, Amann AH. *Topical Drug Delivery Formulations*. New York: Marcel Dekker, 1990.
Yalkowsky SH. *Handbook of Aqueous Solubility Data*. Boca Raton: CRC Press, 2003.
Yalkowsky SH. *Techniques of Solubilization of Drugs*. New York: Marcel Dekker, 1981.

Equipment

Busse DJ. *Mfg Chem* 1990; 61:39.
Lagman B. *Drug Develop Ind Pharm* 1988; 14:2705.
Oldshue JY. *Fluid Mixing Technology*. New York: McGraw-Hill, 1983.

Excipient Properties

Kibbe AH, ed. *Handbook of Pharmaceutical Excipients,* 3rd ed. Washington, DC: American Pharmaceutical Association, 2000.
Reynolds JEF, ed. *Martindale, The Extra Pharmacopoeia*, 31st ed. London: Pharmaceutical Press, 1996.

REFERENCES

1. Wang J, Yang TY, Zhang JX. *Herald of Medicine* 2003; 22:642–643.
2. Schmidt-Nawrot J. *Pharm Ind* 2000; 62:464–469.
3. Eisinger HJ. *Pharm Ind* 2000; 62:469–473.
4. Pfafflin A. *Pharm Ind* 2000; 62:223.
5. Woiwode W, Huber S. *Pharm Ind* 2000; 62:377–381.
6. Connors KA, Amidon GL, Stella VJ. *Chemical Stability of Pharmaceuticals: A Handbook for Pharmacists,* 2nd ed. New York: Wiley-Interscience, 1986.
7. Kurup T, Wan L. *Pharm J* 1986; 37:761.
8. Novack GD, Evans R. *J Glaucoma* 2001; 10:483–486.
9. *United States Pharmacopeia 27 / National Formulary 22*. Rockville, MD: United States Pharmacopeial Convention, Inc., 2004.
10. Sutton SVW, Porter D. *J Pharm Sci Technol* 2002; 56:300–311.
11. Clesceri LS, Greenberg AE, Eaton AD, eds. *Standard Methods for Examination of Water and Wastewater,* 20th ed. Washington DC: American Public Health Association, 1998.
12. Moll F, Naeff REJ, Ehrhart EI, et al. *Pharm Ind* 1997; 59:258–264.
13. Coates D. *Manuf Chem Aerosol News* 1974; 45:19–20.
14. Steinberg DC. 1995; 110: 71–76.
15. Koch CS. *Parfuem Kosmet* 1994; 75:6–21.
16. Bruch CW. *Drug Cosmet Ind* 1976; 118:49–53; 161–162.
17. Ma MH, Lee T, Kwong E. *J Pharm Sci* 2002; 91:1715–1723.
18. Maa YF, Hsu CC. *Int J Pharm* 1996; 140:155–168.
19. Scalzo M, Orlandi C, Simonetti N, et al. *J Pharm Pharmacol* 1996; 48:1201–1205.
20. Coates D. *Manuf Chem Aerosol News* 1973; 44:34–37.
21. Coates D. *Manuf Chem Aerosol News* 1973; 44:41–42.
22. Hurwitz SJ, McCarthy TJ. *J Clin Pharm Ther* 1987; 12:107–115.
23. Boussard P, Devleeschouwer MJ, Dony J. *Int J Pharm* 1991; 72:51–55.
24. Gionchetti P, Venturi A, Rizzello F, et al. *Aliment Pharmacol Ther* 1997; 11:679–684.
25. Matula C, Nahler G, Kreuzig F. *Int J Clin Pharmacol Res* 1988; 8:259–261.
26. Breytenbach HS. *Curr Ther Res-Clin Exp* 1979; 26:640–643.
27. Allen LV. *US Pharmacist* 1990; 15:88–90.
28. Davis CC, Squier CA, Lilly GE. *J Periodont* 1998; 69:620–631.
29. Shibly O, Ciancio SG, Kazmierczak M, et al. *J Clin Dent* 1997; 8:145–149.

30. Hanawa T, Masuda N, Mohri K, et al. *Drug Dev Ind Pharm* 2004; 30:151–161.
31. Amerongen AVN, Veerman ECI. *Support Care Cancer* 2003; 11:226–231.
32. Ellis ME, Clink H, Ernst P, et al. *Eur J Clin Microbiol Infect Dis* 1994;13:3–11.
33. Ellepola ANB, Samaranayake LP. *Oral Dis* 2001; 7:11–17.
34. Joseph BK. *Med Princ Pract* 2002; 11:32–35.
35. Tricca RE. 1988; 142:32.
36. Jungnickel PW, Shaefer MS, Maloley PA, et al. *Ann Pharmacother* 1993; 27:700–703.
37. Abdel-Rahman SM, Johnson FK, Gauthier-Dubois G, et al. *J Clin Pharmacol* 2003; 43:148–153.
38. Krieger JN. *J Urol* 2002; 168:2351–2358.
39. Sarkar MA. *Pharm Res* 1992; 9:1–9.
40. Eppstein DA, Longenecker JP. 1988; 5:99–139.
41. Davis SS, Illum L. *Clin Pharmacokinet* 2003; 42:1107–1128.
42. Arnold J, Ahsan F, Meezan E, et al. *J Pharm Sci* 2002; 91:1707–1714.
43. Bateman ND, Whymark AD, Clifton NJ, et al. *Clin Otolaryngol* 2002; 27:327–330.
44. Tsikoudas A, Homer JJ. *Clin Otolaryngol* 2001; 26:294–297.
45. Adams WP, Conner CH, Barton FE, et al. *Plast Reconstr Surg* 2001; 107:1596–1601.
46. Connolly JG, Anderson C. *Can Med Assoc J* 1979; 121:318–320.
47. Abbas AAH, Felimban SK, Yousef AA, et al. *Med Pediatr Oncol* 2002; 39:139–140.
48. Mallet L, Sesin GP, Ericson J, et al. *N Engl J Med* 1982; 307:445.
49. Greenwood J. *Pharm J* 1989; 243:553–557.
50. Jain NK, Rosenberg DB, Ulahannan MJ, et al. *Am J Gastroenterol* 1985; 80:678–681.
51. Kim NC, Kinghorn AD. *Arch Pharm Res* 2002; 25:725–746.
52. Mitchell JC, Counselman FL. *Acad Emerg Med* 2003; 10:400–403.
53. Cooper C, Kilham H, Ryan M. *Med J Aust* 1998; 168:94–95.
54. Cote CJ, Cohen IT, Suresh S, et al. *Anesth Analg* 2002; 94:37–43.
55. Allen LV, Erickson MA. *Am J Health-Syst Pharm* 1998; 55:1915–1920.
56. Allen LV, Erickson MA. *Am J Health-Syst Pharm* 1998; 55:1804–1809.
57. Horner RK, Johnson CE. *Am J Hosp Pharm* 1991; 48:293–295.
58. Al-Waili NS. *Complement Ther Med* 2003; 11:226–234.
59. Nussinovitch A. *Water-Soluble Polymer Applications inFfoods*. Oxford: Blackwell Science, 2003.
60. Bedinghaus JM, Niedfeldt MW. *Am Fam Physician* 2001; 64:791–796.
61. Kim CK, Yoon YS, Kong JY. *Int J Pharm* 1995; 120:21–31.
62. Williams RO, Rogers TL, Liu J. *Drug Dev Ind Pharm* 1999; 25:1227–1234.
63. Smith KJ, Chan HK, Brown KF. *J Aerosol Med-Depos Clear Eff Lung* 1998; 11:231–245.
64. Weisshaar E, Dunker N, Gollnick H. *Neurosci Lett* 2003; 345:192–194.
65. Stozek T, Borysiewicz J. *Pharmazie* 1991; 46:39–41.
66. Park SJ, Kim SH. *J Colloid Interface Sci* 2004; 271:336–341.
67. Rubinstein A, Pathak YV, Kleinstern J, et al. *J Pharm Sci* 1991; 80:643–647.
68. Constantinides PP. *Pharm Res* 1995; 12:1561–1572.
69. Bhargava HN, Narurkar A, Lieb LM. *Pharm Tech* 1987; 11:46.
70. Griffin WC, Lynch MJ, Lathrop LB. *Drug Cosmet Ind* 1967; 101:41.
71. Nielloud F, Marti-Mestres G, eds. *Pharmaceutical Emulsions and Suspensions*. New York: Marcel Dekker, 2000.
72. Lachman L, Liebermann HA, Kanig J, eds. *The Theory and Practice of Industrial Pharmacy,* 3rd ed. Philadelphia: Lea & Febiger, 1986.
73. Binks BP. *Curr Opin Colloid Interface Sci* 2002; 7:21–41.
74. Florence AT, Whitehill D. *Int J Pharm* 1982; 11:277–308.
75. Okochi H, Nakano M. *Adv Drug Deliv Rev* 2000; 45:5–26.
76. Kassem MA, Safwat SM, Attia MA, et al. *STP Pharma Sci* 1995; 5:309–315.
77. Bourrel M, Schechter RS, eds. *Microemulsions and Related Systems: Formulation, Solvency, and Physical Properties*. New York: Marcel Dekker, 1988.
78. Rosano HL, Cavallo JL, Chang DL, et al. 1988; 39:201–209.
79. Lin TJ, Shen YF. *J Soc Cosmet Chem* 1984; 35:357–368.
80. Lin TJ. *J Soc Cosmet Chem* 1978; 29:117–125.
81. Maa YF, Hsu C. *J Control Release* 1996; 38:219–228.
82. Lieberman HA, Rieger MM, Banker GS, eds. *Pharmaceutical Dosage Forms. Volume 3: Disperse Systems,* 2nd ed. New York: Marcel Dekker, 1998.
83. Maa YF, Hsu CC. *Pharm Dev Technol* 1999; 4:233–240.
84. van Balen GP, Martinet CAM, Caron G, et al. *Med Res Rev* 2004; 24:299–324.
85. Derycke ASL, de Witte PAM. *Adv Drug Deliv Rev* 2004; 56:17–30.
86. Gregoriadis G, Florence AT, Patel HM, eds. *Liposomes in Drug Delivery*. Vol 2. Drug Targeting and Delivery. Langhorne, PA: Harwood Academic Publishers, 1993.
87. Lasic DD, Papahadjopoulos D, eds. *Medical Applications of Liposomes*. Amsterdam: Elsevier, 1998.
88. Bangham AD, Standish MM, Watkins JC. 1965; 13:238.
89. *Physicians' Desk Reference,* 58th ed. Montvale, NJ: Medical Economics, 2003.
90. Tsai SC, Botts D, Plouff J. *J Rheol* 1992; 36:1291–1305.
91. FDA. *Guidance for Industry. Stability Testing of Drug Substances and Drug Products*. Center for Drug Evaluation and Research (CDER). 1998; 1–114.
92. Levinson ML, Johnson CE. *Am J Hosp Pharm* 1992; 49:122–125.
93. Harris AM, Rauch AM. *Pediatr Infect Dis J* 1994; 13:838–838.
94. Echizen H, Ishizaki T. *Clin Pharmacokinet* 1991; 21:178–194.
95. Tortorici MP. *Am J Hosp Pharm* 1979; 36:22.
96. Morales ME, Lara VG, Calpena AC, et al. *J Control Release* 2004; 95:75–81.
97. Shah KP, Chafetz L. *Int J Pharm* 1994; 109:271–281.
98. Sjoqvist R, Graffner C, Ekman I, et al. *Pharm Res* 1993; 10:1020–1026.
99. Sheumaker JL. United States Patent #4,762,709. 1988.
100. Tipton AJ, Holl RJ. United States Patent #5,747,058. 1998.
101. Nail SL, White JL, Hem SL. *J Pharm Sci* 1976; 65:1195–1198.
102. Moore WN, Taylor LT. *J Nat Prod* 1996; 59:690–693.

Sterilization

Donald E Hagman PhD

Sterilization is an essential concept in the preparation of sterile pharmaceutical products. Its aim is to provide a product that is safe and eliminates the possibility of introducing infection.

Mergers and acquisitions of pharmaceutical companies create multinational organizations faced with complying with all of the regulatory agencies of the involved countries. To date there is no global regulatory agency that oversees the production of sterile pharmaceutical products. Multinational companies must be familiar with the regulations of all countries in which they operate and meet those regulations. Although it is not the intent of this chapter to delineate the sterilization standards for all countries, it is to provide a detailed description of the techniques used throughout the world to sterilize pharmaceutical products. There are many attempts to standardize practices throughout the multinational industry. These include the efforts of the International Council on Harmonization (ICH) and the issuance of various technical ISO standards and compendial efforts of the various countries like United States Pharmacopoeias (USP) to set some basic standards. Additionally, organizations like International Society for Pharmaceutical Engineering (ISPE) and Parenteral Drug Association (PDA) have issued various documents, which include all facets of the international regulatory requirements.

Sterilization is a process used to destroy or eliminate viable microorganisms that may be present in or on a particular product or package. The process requires an overall understanding and control of all parts of the preparation for use of a particular product. Those areas include the selection and acceptance of all materials used for the product and package, environment in which the product is prepared and used and the ultimate disposition of the remaining materials after use. Sterilization may be required for several steps of the process using any one or a combination of the techniques listed in this chapter.

The aim of a sterilization process is to destroy or eliminate microorganisms that are present on or in an object or preparation, to make sure that this has been achieved with an extremely high level of probability and to ensure that the object or preparation is free from infection hazards. The currently accepted performance target for a sterilization process is that it provide for a probability of finding a nonsterile unit of less than 1 in 1 million. That is, the process (including production, storage, and shipment) will provide a *Sterility Assurance Level* (SAL) equal to or better than 10^{-6}. This is achieved through the processing of products in validated equipment and systems. Thorough validation and periodic requalification is essential to meeting these sterility requirements.

The purpose of this chapter is to provide a basic understanding of the following sterilization methods currently being used in pharmaceutical technology and the equipment employed to carry out these methods:

Method	Equipment
Moist heat sterilization	Saturated steam autoclaves
	Superheated water autoclaves
	Air over steam autoclaves
Dry heat sterilization	Batch sterilizers
	Continuous tunnel sterilizers
Chemical *cold* sterilization	Ethylene oxide
	Vaporized hydrogen peroxide
	Hydrogen peroxide/steam
	Other gases
Radiation sterilization	Electromagnetic
	Particulate
Filtration	Membranes

DEFINITIONS

The following terms, relating to sterilization, should be understood by those carrying out sterilization processes or handling sterile products:

Antiseptic—A substance that arrests or prevents the growth of microorganisms by inhibiting their activity without necessarily destroying them.

Aseptic—Refers to areas and practices where the intent is to be sterile.

Aseptic Processing—Those operations performed between the sterilization of an object or preparation and the final sealing of its package. These operations are, by definition, carried out in the complete absence of microorganisms.

Bactericide—Any agent that destroys microorganisms.

Bacteriostat—Any agent that arrests or retards the growth of microorganisms.

Bioburden—The number of viable microorganisms present prior to sterilization; Usually expressed in colony-forming units of volume.

Disinfection—A process that decreases the probability of infection by destroying vegetative microorganisms, but not ordinarily bacterial spores. The term usually is applied to the use of chemical agents on inanimate objects.

Germicide—An agent that destroys microorganisms, but not necessarily bacterial spores.

Sanitization—A process that reduces the level of bioburden in or on a product or object to a safe level.

Sterile—The absolute absence of viable microorganisms. There is no degree or partiality.

Sterility Assurance Level (SAL)—An estimate of the effectiveness of a sterilization process. It usually is expressed in terms of the negative power of 10 (ie, 1 in 1 million = 10^{-6}).

Sterilization—A process by which all viable microorganisms are removed or destroyed, based on a probability function.

Terminal Sterilization—A process used to render products sterile to a preferred SAL.

Validation—The act of verifying that a procedure is capable of producing the intended result under prescribed circumstances and challenges to predefined specifications.

Viricide—An agent that will destroy viruses.

STERILITY AS A TOTAL SYSTEM

It is necessary to reiterate the concept already briefly addressed in the introduction. The task of the technology we are dealing with is to provide the product in sterile conditions to the end user. It is currently acknowledged that the quality of the product must be *built into* the process. This concept is particularly true when one of the essential qualities of the product is sterility.

Accordingly, the above-mentioned task is accomplished with a series of design, production, and distribution steps that can be summarized as activities for the selection and routine checking of the following items:

- Active constituents, additives, raw materials in general
- Water used both as solvent and as washing/rinsing agent
- Packaging suitable for the product and for the sterilization process that will be used
- Working environment and equipment
- Personnel

These procedures clearly have the purpose of providing the sterilization process with a product that has a minimum, definite, and consistent bioburden. There are also the following activities:

- Selection of the sterilization method that most suits the unit formed by the product and its packaging, and definition of the process variables for obtaining the intended SAL
- Selection of the machine that is most suitable for performing the selected method and of the utilities that this machine requires
- Qualification and validation of the machine and of the process
- Routine checking of the process
- Checking of the results of the sterilization process
- Proper storage of sterile goods and verification that their sterility is maintained with full reliability throughout the allowed storage period
- Delivering, opening, and using sterile goods without recontamination.

It also should be noted that, in December 2002, the US Food and Drug Administration (FDA) proposed new regulations for aseptic processing and terminal sterilization. The proposed rules as defined in their Concept Paper require that manufacturers of sterile products use validated and robust sterilization techniques wherever possible. The European Pharmacopeia and related pharmacopeias have modified their requirements in their rulings identified as Annex 1.

CONTAMINATION

Certain facts about microorganisms must be kept in mind when preparing sterile products. Some microbes (bacteria, molds, etc) multiply in the refrigerator, others at temperatures as high as 60°C. Microbes vary in their oxygen requirements from the strict anaerobes that cannot tolerate oxygen to aerobes that demand it. Slightly alkaline growth media will support the multiplication of many microorganisms while others flourish in acidic environments. Some microorganisms have the ability to use nitrogen and carbon dioxide from the air and thus can actually multiply in distilled water. In general, however, most pathogenic bacteria have rather selective cultural requirements, with optimum temperatures of 30° to 37°C and a pH of 7.0. Contaminating yeasts and molds can develop readily in glucose and other sugar solutions.

Actively growing microbes are, for the most part, vegetative forms with little resistance to heat and disinfectants. However, some forms of bacteria—among them the bacteria that cause anthrax, tetanus, and gas gangrene—have the ability to assume a spore state that is very resistant to heat as well as to many disinfectants. For this reason, an excellent measure of successful sterilization is whether the highly resistant spore forms of nonpathogenic bacteria have been killed.

The nature of expected contamination and the bioburden are important to pharmacists preparing materials to be sterilized. The raw materials they work with rarely will be sterile, and improper storage may increase the microbial content. Because the pharmacist seldom handles all raw materials in a sterile or protected environment, the environmental elements of the manufacturing area (air, surfaces, water, etc) can be expected to contribute to the contamination of a preparation. The container or packaging material may or may not be presterilized and thus may contribute to the total microbial load.

Understanding the nature of contaminants prior to sterilization and application of methods for minimizing such contamination is vital to preparing for successful pharmaceutical sterilization. Examples of such methods include:

- Maintenance of a hygienic laboratory
- Frequent disinfection of floors and surfaces
- Minimization of traffic in and out of the area
- Refrigerated storage of raw materials and preparations that support microbial growth
- Use of laminar airflow devices for certain critical operations
- Use of water that is of appropriate USP quality and is free of microbial contamination (It is preferable to use presterilized water to avoid any possible contamination.)

METHODS

General

The procedure to be used for sterilizing a drug, a pharmaceutical preparation, or a medical device is determined to a large extent by the nature of the product. It is important to remember that the same sterilization technique cannot be applied universally because the unique properties of some materials may result in their destruction or modification. Methods of inactivating microorganisms may be classified as either physical or chemical. Physical methods include moist heat, dry heat, and irradiation. Sterile filtration is another process, but it only removes, not inactivates, microorganisms. Chemical methods include the use of either gaseous or liquid sterilants. Guidelines for the use of many types of industrial and hospital sterilization are available.[1–10]

Each sterilization method can be evaluated using experimentally derived values representing the general inactivation rates of the process. For example, a death rate or survival curve for a standardized species can be diagramed for different sterilization conditions. This is done by plotting the logarithm of surviving organisms against time of exposure to the sterilization method. In most instances, these data show a linear relationship, typical of first-order kinetics, and suggest that a constant proportion of a contaminant population is inactivated in any given time interval. Based on such inactivation curves, it is possible to derive values that represent the general inactivation rates of the process. For example, based on such data, it has become common to derive a decimal reduction time or D value, which represents the time under a stated set of sterilization exposure conditions required to reduce a surviving microbial population by a factor of 90%.

D values, or other expressions of sterilization process rates, provide a means of establishing dependable sterilization cycles. Obviously, the initial microbial load on a product to be sterilized becomes an important consideration. Beyond this, however, kinetic data also can be used to provide a statistical basis for the success of sterilization cycles. A simple example will
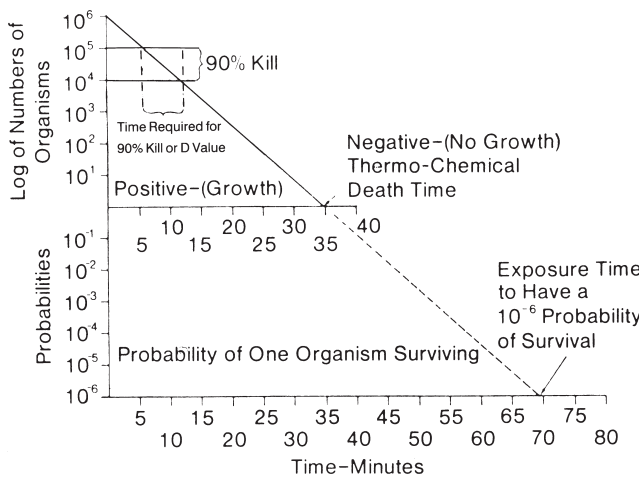
Figure 40-1. Sterilization model using D values.

suffice (Fig 40-1). When the initial microbial contamination level is assumed to be 10^6, and if the D value of the sterilization process is 7 minutes, complete kill is approached by application of 6 D values (42 minutes). However, at this point reliable sterilization would not be assured because a few abnormally resistant members of the population may remain. In this example, by extending the process to include an additional 6 D values, most of the remaining population is inactivated, reducing the probability of one organism surviving to one in 1 million.

Moist Heat

ESSENTIALS OF STEAM STERILIZATION KINETICS

Let us suppose a system contaminated by microorganisms (which we assume, for the sake of simplicity, to be pure and homogeneous) is immersed in pressurized saturated steam, at constant temperature; for example, it could be a vial containing an aqueous suspension of a certain spore-forming microorganism.

It has been shown experimentally that, under the above conditions, the reaction of thermal degradation of the microorganism obeys the laws of chemical reactions: the rate of reduction of the number of microorganisms present in the system in each moment is proportional to the actual number itself. The proportionality coefficient is typical of the species and conditions of the chosen microorganism.

Thus, the degradation reaction (the sterilization process) develops like a first-order chemical reaction in which the reaction rate is proportional, in each moment, only to the amount of microorganisms still to be inactivated. This seems to be obvious for dry sterilization, but less rigorous for steam sterilization, in which the water vapor molecules also seem to take part in the reaction. Actually, this bimolecular reaction is of the first order, as the steam is present in high excess during the entire reaction and its concentration may be regarded as constant.

The most frequently used mathematical expression of the above facts is

$$N = N_0 \, 10^{-t/D} \qquad (1)$$

where N_0 is the initial number of microorganisms, t is the elapsed exposure (equal to sterilization time), N is the number of microorganisms after the exposure time t, and D is the *decimal decay time,* defined as the time interval required, at a specified *constant* temperature, to reduce the microbial population being considered by 1/10 (ie, by one logarithmic value; eg, from 100% to 10% or from 10% to 1% of the initial value).

The D value is inversely proportional to the first-order reaction coefficient and is therefore typical of the species and condi-

tions of the chosen microorganism. Depending on the initial hypothesis of exposure at constant temperature, each D value always refers to a specified temperature.

Equation 1 allows one to draw a first very important conclusion: the time required to reduce the microorganism concentration to any preset value is the function of the initial concentration. The sterilization reaction is therefore neither an *all-or-nothing* process nor a *potential barrier* process as was once thought.

It also is evident immediately that the effect of sterilization at the same constant temperature will be very different depending on the D value of the contaminating microbial species (or on the largest D value in the usual case of mixed contamination). Figure 40-2 shows that the same reduction ratio for different species is achieved after exposure time proportional to the D value of each species. The graph derives only from Equation 1 and from the definition of D value. The basic hypothesis of the temperature being constant is thoroughly valid.

Sterility Is a Probable Effect of Exposure Time—Let us now consider what happens within a batch of units (vials, bottles, or others) with an initial constant unit contamination of 100 microorganisms equal to 10^2. If the D value at 121°C is assumed to be 1, after 1 min at 121°C, a reduction equal to 10^1 = 10 microorganisms is achieved; after another minute, only 10^0 = 1 microorganism is still surviving. After another minute, the surviving microbial population would be 10^{-1} = 1/10 microorganism. A contamination of 1/10 must not be understood to mean that each unit contains 1/10 of a microorganism, which is biologically meaningless (in this case the unit probably would be sterile) but that there is a probability of having 1/10 of the units still contaminated within the batch of sterilized units.

In fact, 3 min would be the necessary time to reduce the microbial population to a single surviving microorganism if the initial population were 10 times larger than the one at issue. This higher initial contamination could be regarded either as a 10 times larger number of microorganisms in the same unit, or as the initial contamination of a 10 times larger unit.

If the unit is not considered any longer as the single vial or bottle, but as the whole of all the items produced over a period of time, the initial number of microorganisms present in each item has to be multiplied times the number of items produced, and the exposure time to achieve the reduction to the same number of viable microorganisms left in the whole of the items produced, has to be increased correspondingly. The following example will be helpful to focus the matter.

A new sterile product in ampules has to be manufactured; the number of ampules to be produced over all the life period of the product is expected to be 10^{10}. The maximum number of contaminated ampules deemed to be acceptable is 10 = 1: this obviously means that the probability of having nonsterile ampules after sterilization must not exceed 10^{-10}. Let us also suppose that the microbial population within each ampule after the filling and the sealing does not exceed 10^3 microorganisms. These must be destroyed by means of moist heat-terminal sterilization at 121°C. The applicable D value is 1 min. The total number of microorganisms to be destroyed during the life of the product will be

$$10^{10+3} = 10^{13}$$

If this whole microbial population were exposed to moist heat at 121°C over a period of 13 min, it would be reduced to 10^{-13} times its initial number (ie, to $10^{13-13} = 10^0 = 1$. The exposure time of 13 min thus would be sufficient (under all the other above hypotheses) to prevent the total number of contaminated ampules from exceeding the value of 1.

From the point of view of each single ampule, 13 min of exposure would reduce the microbial population to the theoretical value of

$$10^{3-13} = 10^{-10}$$

To interpret this numeric value as the probability of still having one contaminated ampule in 10 billion sterilized ampules means that a single ampule will still be contaminated out of a whole lot of 10^{10}. This probability value is defined as PNSU (probability of nonsterile unit).

In recent times the PNSU as a sterility evaluation criterion is being replaced by the SAL. The name itself could generate some misunderstanding, because a level of assurance commonly is deemed to be good if high, but SAL seems to have been defined in such a way that its numerical value is the same as PNSU. This notwithstanding, it is sometimes calculated as the reciprocal value of PNSU. The SAP (sterility assurance probability) criterion has been proposed as well and SAP seems for the moment to have been granted the same definition of PNSU, even if it would be better understandable if its value approached unity after a satisfactory sterilization.

The above discussion and example lead to the conclusion that the optimum exposure time for a sterilization process must take into account not only the initial microbial population within the single item to be sterilized and the species and conditions of the contaminating microorganism, but also the total number of items expected to be sterilized over the life of the product.

Effect of Temperature Changes—All the above considerations have been developed under the basic assumption that the temperature is kept constant during the entire exposure time. It seems rather obvious that the D value will change as the temperature changes. If the D values experimentally obtained for a given microbial species are plotted on a semilogarithmic chart as the function of the temperature T, a path similar to Figure 40-3 is obtained.

In this case, it can be seen that D value is 1 min at 121°C (ie, the average value which very often is assumed to be acceptable in the absence of more exact experimental data). It also can be seen that D value varies by a factor of 10 if the temperature varies by 10°C.

The z value is defined as the temperature coefficient of microbial destruction, the number of degrees of temperature that causes a 10-fold variation of D (or, more generally, of the sterilization rate). The z values generally oscillate between 6 and 13 for steam sterilization in the range 100° to 130°C, and z value often is assumed to be equal to 10 in the absence of more precise experimental data.

The fact that D value varies by 10 times for a variation of 10°C when $z = 10$ must not lead to the false assumption that D varies by one time (ie, doubles) for an increase of 1°C. Obviously, this is not true. It is actually a matter of finding the number which yields 10 when raised to the tenth power. This number is 1.24. Therefore, a variation of 1°C entails a variation of D value of 24%. This is quite a significant number, which illustrates the dramatic effects that are generated when the sterilization temperature is also only a few degrees lower than the expected value, perhaps only in some areas of the sterilizer load.

It is also useful to remember that the effect of temperature variation decreases considerably as the temperature rises and drops to approximately 1/2 (or even less) for dry sterilization at approximately 200°C. Under these conditions the z value is about 20 instead of about 10. Therefore, the small temperature differences that can be so dramatic in steam sterilization have much less effect in dry sterilization.

The foregoing refers to average values because the actual D values and z values depend to a large extent on the medium that contains the microorganisms and on their history. At 121°C no microorganism has exactly D = 1 and $z = 10$. However, the combined use of these two parameters in calculating F_0 and PNSU provides ample margins of safety with regard to the microorganisms with which we deal commonly.
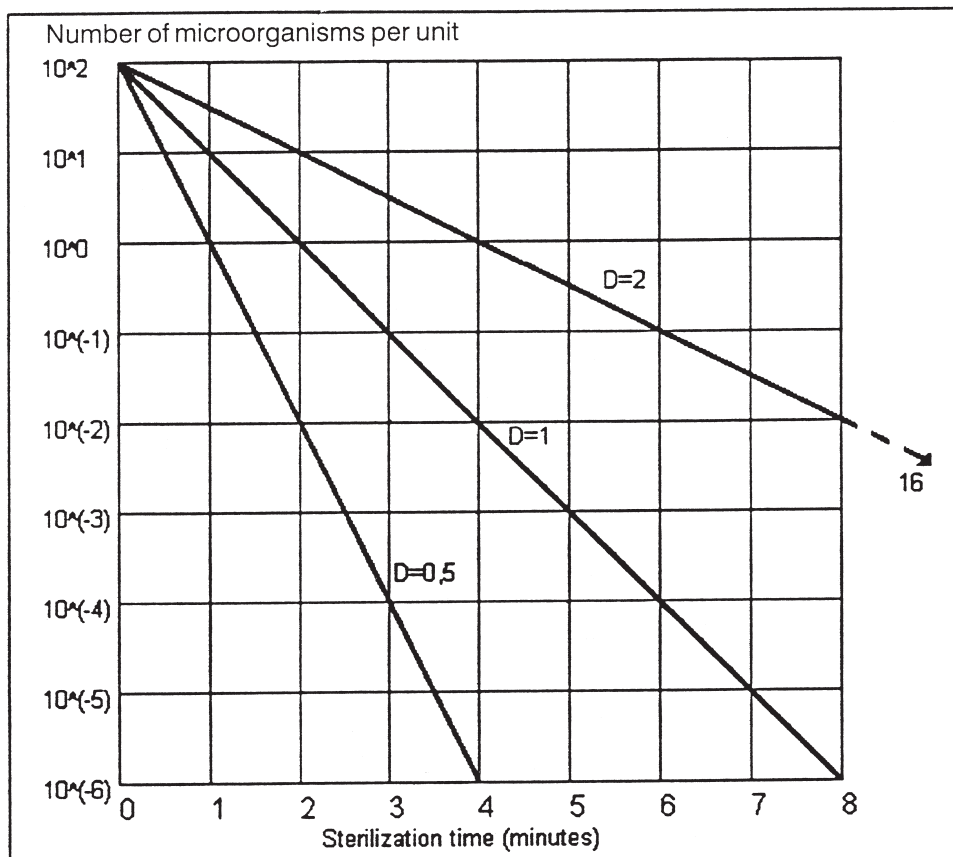

Figure 40-2. Effect of varying D values on sterilization rate (courtesy, Fedegari Autoclavi).
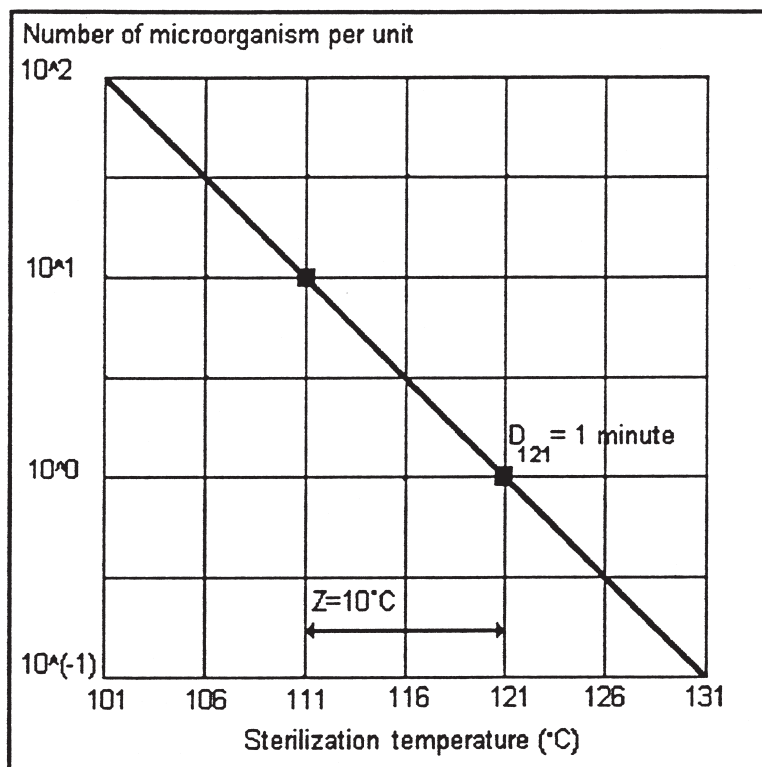
Figure 40-3. Effect of temperature on microbial destruction (courtesy, Fedegari Autoclavi).

F_0 *or Equivalent Sterilization Time at 121°C*—It is of the utmost interest to calculate the lethal effect of the exposure of a microbial population to a variable temperature, T, by relating it to an hypothetical sterilization performed at a constant temperature, T_0, for the time, t_0. If the constant reference temperature is assumed equal to 121.1°C (originally 250°F) and the z value equal to 10, the equivalent time is termed F_0. Thus, F_0 is the equivalent exposure time at 121.1°C of the actual exposure time at a variable temperature, calculated for an ideal microorganism with a temperature coefficient of destruction equal to 10.

First introduced in the *Laboratory Manual for Food Canners and Processors* by the National Canners Association in 1968, F_0 has become a common term in pharmaceutical production since the FDA used it extensively in the "Proposed Rules" of June 1, 1976 (21 CFR 212.3) with the following meaning:

F_0 means the equivalent amount of time, in minutes at 121.1°C (250°F), which has been delivered to a product by the sterilization process. For the calculation of it, a z value of 10°C or 18°F is assumed; the term z value means the slope of the thermal death time curve and may be expressed as the number of degrees required to bring about a 10-fold change in the death rate.

In practice, the knowledge of the temperature values as the continuous function of elapsing time is not available, and F_0 is calculated as

$$F_0 = \Delta t \, \Sigma \, 10^{\frac{T-121.1}{z}} \qquad (2)$$

where Δt is the time interval between two consecutive measurements of T, T is the temperature of the sterilized product at time t, and z is the temperature coefficient, assumed to be equal to 10.

Saturated Steam

PRINCIPLES

Sterilization with saturated steam is the method that provides the best combination of flexibility in operation, safe results and low plant and running costs. The sterilizing medium obviously is pressurized saturated steam and the typical operating temperature is 121°C (250°F), but higher or lower temperatures often are used.

The term *dry* saturated steam sometimes is used: it should be made clear that this is an *ideal* condition of steam, and that moist saturated steam is used in practice for sterilization. However, the steam must entrain the smallest possible amount of condensate. The *water vapor ratio* of the steam defines the amount of condensate entrained by 100 parts by weight of moist steam; a water vapor ratio of 0.95 means that 100 g of steam consist of 95 g of dry saturated steam plus 5 g of condensate which is, or should be, *at the same temperature as the steam.*

The reliability of sterilization performed with saturated steam is based on several particular characteristics of this medium. When steam condenses, it releases calories at a constant temperature and in a *considerable amount:* 1 kg of pure saturated steam condensing at 121°C (turning into water at 121°C, thus without cooling) releases as much as 525 kcal. The temperatures and pressures of saturated steam have a two-way correlation. Once the temperature of the steam is determined, so is its pressure, and *vice versa.* Saturated steam at 121°C inevitably has a pressure of 2.05 abs bar. This entails two very interesting practical possibilities:

1. A pure saturated steam autoclave can be controlled indifferently according to the temperature parameter or according to the pressure parameter.
2. Regardless of the parameter used for control, the second parameter can be used easily to cross-monitor the first one.

A 1 gram molecule of water (18 g, or 18 mL in the liquid state) as steam at 121°C and 2.05 abs bar occupies a volume of approximately 15 L. This means that when steam condenses at 121°C it shrinks in volume by almost 1000 times. Accordingly, additional available steam *spontaneously* reaches the object to be sterilized. The condensate that forms can be removed easily from the autoclave chamber by means of a condensate discharge or, with a more modern technique, by continuous and

forced bleeding (as occurs for example in so-called *dynamic steam* sterilizers).

However, several other phenomena must be considered. To perform its microorganism inactivating action (coagulation of cellular proteins), the steam, or more generally the moist heat, must make contact with the microorganisms. This can occur directly or indirectly. For example, it occurs directly when the steam that is present in the autoclave chamber is in direct contact with a surgical instrument. It instead occurs indirectly when moist steam is generated (by heat exchange with the steam present in the chamber) inside a sealed ampule that contains an aqueous solution. However, it is evident that it is not possible to steam-sterilize the inside of an empty closed ampule or the contents of an ampule if they are constituted by an anhydrous oil-based solution.

The air that is initially present in the autoclave chamber and the *incondensables* that possibly are entrained by the steam (generally CO_2) have molecular weights, and thus densities, 1.5 to 2.0 times higher than steam (under equal temperature/pressure conditions). Therefore, the air must be eliminated initially from the chamber and the steam must not introduce incondensables in the chamber; otherwise, these tend to stratify in the lower portions of the chamber, creating intolerable temperature gradients.

When closed nondeformable containers that contain aqueous solutions are sterilized, the pressures inside them can reach values far above those of the chamber. All air has been removed from the chamber, which in fact only contains steam: accordingly, at 121°C the pressure is 2.05 abs bar. The container instead almost always has a head space that contains air (or other gases). During sterilization, the aqueous solution of the container produces a vapor pressure that is approximately equal to 2.05 abs bar, but this value is increased by the partial pressure of the air of the head space; assuming that its initial value is 1.0 bar, it will increase to approximately 1.3 bar due to heating.

Pressure increases also will occur due to the thermal expansion of the solution (which is not entirely compensated by the expansion of the glass of the container) and because any gases dissolved in the solution may leave it.

Generally, in the conditions described above the total pressure inside the container exceeds by approximately 1.4 bar the pressure in the chamber if the initial head space is, as usually occurs, 10 to 20% of the total volume of the container. This overpressure generally is well tolerated by glass ampules, even those of considerable capacity (20 to 30 mL). However, it becomes hazardous for glass containers fitted with rubber stoppers held in place by a seal (due to the risk of stopper lifting) and intolerable for deformable containers, such as rigid (and even flexible) plastic containers, prefilled syringes, or cans. In all these cases, it is necessary or convenient to use the *counterpressure* sterilization methods (described later).

SATURATED STEAM AUTOCLAVES

Materials—All autoclaves intended for the pharmaceutical industry are made of Class AISI 316 stainless steel, including valves and piping (Fig 40-4). Only the service elements arranged *downstream* of the autoclave (for example the vacuum pump or the condensate discharge) are accepted if they are made of other materials. The service elements *upstream* of the autoclave (eg, heat exchangers or water pumps) also must be made of stainless steel.

Silicone rubber or Teflon and derivatives thereof generally are used for the gaskets (of doors, valves, etc).

Structure—Saturated steam autoclaves generally have a quadrangular, or rarely cylindrical, chamber. The doors are generally quadrangular even if the structure is cylindrical; in this case, the doors are inscribed in the circumference. There may be one or two doors: when the autoclave leads to a sterile room, there are always two doors.

Two-door autoclaves often are used when this requirement does not occur but the need is nonetheless felt to separate the

Figure 40-4. A modern computerized steam autoclave with horizontal sliding door (courtesy, Fedegari Autoclavi).

loading area, where products to be sterilized are placed, from the unloading area, where already sterilized products are placed. *This concept applies to all types of sterilizers.*

Doors may be of various kinds. The most common types are

- Hinged, manually operated, retained by radial locking bars, with a solid and fixed gasket
- Hinged, semiautomatically operated, retained by means of abutments in which the door engages automatically and with a movable gasket activated by compressed air
- Vertically or laterally sliding, with retention and gaskets as mentioned immediately above

Saturated steam autoclaves generally are jacketed. There is no room here to discuss the various kinds of jacket and their purposes. However, there are two ways to feed steam into the jacket and into the chamber:

Single Feed—the steam circulates first in the jacket and passes from the jacket into the chamber.
Separate Feed—usually the chamber is fed pure steam and the jacket is fed industrial steam.

Single-feed steam has some advantages in terms of control, but separate-feed steam is preferred because it provides better assurances of lack of microbiological and particle contamination.

MANAGEMENT SYSTEMS

The management systems used on currently manufactured autoclaves are programmable logic controllers (PLCs) or personal computers (PCs), or sometimes combinations of PLCs and PCs. This is also true for other kinds of autoclaves and sterilizers, which will be discussed later. However, a very large number of autoclaves controlled by electropneumatic systems are still in operation and still perform acceptable work. Naturally, the current control systems offer a kind of performance that was undreamed of earlier.

Pressure or temperature control (as mentioned previously, these parameters are interchangeable for a saturated steam autoclave) generally is performed with a proportional-integral-derivative method. Control by temperature is the generally accepted scheme because it is not influenced by trapped air. Sterilization can be time-managed or F_0-managed (with the F_0 being accumulated by heat probes enabled for this function), or time-managed with simultaneous calculation of F_0 for monitoring purposes.

Some management systems offer exceptional flexibility in composing programs and in setting parameters even to operators who have no knowledge of electronic programming. The information provided in real time (on same display device) is ex-

tremely detailed, as is the permanent information, which can be produced on paper or stored on various kinds of electronic medium.

PROCESS

Initial Removal of the Air from the Chamber—The main reason the air must be removed from the autoclave chamber has been pointed out above.

Loads often are made up of porous materials or materials packaged in sterilization paper or in plastic/paper bags, or contained in filter boxes. All these situations require reliable and rapid removal of the air from the load. The so-called *gravity* removal method is considered obsolete. Modern autoclaves have a water-ring vacuum pump that can produce a vacuum of approximately 70 residual mbar in the chamber. Accordingly, only about 10% of the air remains in the chamber. There are essentially two methods for completing air removal:

Pulsed Vacuum—Once the initial vacuum has been reached, the pump is stopped and steam is introduced in the chamber (up to approximately atmospheric pressure), then vacuum is produced again. Three or more of these vacuum/steam pulses are performed.

Dynamic Vacuum—Once the initial vacuum has been reached, the pump continues to run, but at the same time a 5- to 10-min injection of steam is performed (from the side of the chamber that lies opposite the vacuum drain).

Modern autoclaves are capable of performing either of these methods, chosen according to the load to be processed.

Heating-Sterilization—During heating phases, and much less during the sterilization phase, considerable amounts of condensate form in the chamber. Except for particular instances, this condensate must be removed from the chamber. There are basically two extraction methods:

- *A condensate trap located at the bottom of the chamber.* This is the simplest and cheapest method, but it causes significant pressure drops, and therefore temperature drops, inside the chamber due to the inertia of the condensate trap. Essentially, it discharges not only the condensate but also significant amounts of steam, which cause instantaneous expansion, and thus cooling, of the steam that remains in the chamber.
- *Dynamic steam.* This is the most reliable and elegant system, but is also more expensive. During the heating and sterilization phases, the vacuum pump is kept running and draws from the chamber all the condensate that forms in it through a low-capacity valve. A certain amount of steam is naturally aspirated continuously, and a dynamic condition of the steam is thus produced, hence the name of the method.

Autoclaves also are required to have a continuous steam bleed past the controlling sensor in the drain line.

Post-Sterilization Phases—These may be different according to the material to be sterilized and depending on the results to be obtained on the material itself. The most common solutions are listed below.

1. **Vacuum and Time-Controlled Vacuum Maintenance**—This method is used to dry and simultaneously cool loads of solid materials, both porous and nonporous. It is performed by restarting the vacuum pump until a preset value (eg, 100 mbar) is reached; the pump then is kept running for a preset time (eg, 20 min).
2. **Cooling by Circulating Cold Water in the Jacket**—This method is used to cool containers that are partially filled with solution (eg, culture media) and closed with sleeve (Bellco-type) stoppers. Naturally, with these loads Item 1 is not applicable, because the solution would boil, and Item 3 is dangerous due to possible contaminations. This method is performed by removing the steam present in the chamber through the introduction of compressed sterile air at a pressure that is equal to, or greater than, the sterilization pressure. Then, cold water is circulated in the jacket. The pressurized compressed air in the chamber has two purposes: (1) to prevent the solution from boiling and (2) to improve heat exchange between the load and the jacket.
3. **Cooling by Spraying Water on the Load**—This method generally is used for loads of filled and closed ampules and plastic intravenous containers. It is performed with deionized water (to avoid salt residues on the ampules) which is nebulized onto the

load by means of a sparger provided in the ceiling of the chamber. Naturally the ampules, which preferably are arranged in an orderly fashion, must be contained in trays with a perforated bottom. Nebulization of the water causes a rapid condensation of the steam that produces a sudden pressure drop in the chamber, whereas the pressure inside the ampules still remains rather high because the solution cools rather slowly. Ampules of good quality (even large ones up to approximately 20 mL) tolerate this method adequately. Cooling stops when the solution inside the ampules has reached the temperature of 70° to 80°C. In this manner, the load, removed from the autoclave, still contains enough heat energy to dry spontaneously.

4. **Ampule Tightness with Fast Vacuum**—The pressure stress described in Item 3, above, is produced deliberately and increased by activating the vacuum pump as soon as the sterilization phase ends. The pressure in the chamber quickly drops to values that can reach 150 to 200 mbar (obviously this value can be controlled easily), whereas the pressure inside the closed ampules initially remains above 3.0 bar. The ΔP thus produced breaks ampules with *closed defects,* such as thinner regions, tensions in the glass, and closed cracks.

   Obviously, if the ampules have *open defects* (ie, holes at the tip or open cracks), the ΔP does not arise or is very small and thus the ampules rarely break. What happens instead is that the solution in the ampule boils and thus evaporates, reducing the volume of the solution. Unfortunately, this evaporation is very limited. Because it requires a considerable amount of energy, the solution cools very quickly and the boiling ends. One cannot rely on the transmission of heat from the adjacent ampules or from the jacket, because the chamber is evacuated. It is evident that in such conditions, solution in the liquid state leaks from the ampules; at least from the *open defects* that lie below the level of the solution. Accordingly, it may be convenient to load the ampules upside down (ie, with their tip pointing downward) if it is known that most defects occur at the tip or shoulder of the ampules. Naturally, the breakage of the ampules or the leakage of solution soils the load, which must therefore be washed and dried. With appropriate methods it usually is possible to achieve all this in the autoclave itself.

5. **Cooling as in Item 3, but with Air Counterpressure**—In many cases it is not possible or reasonable to subject the load, during cooling, to the pressure stress that arises with the method described in Item 3. In such cases, it is possible to remove the steam present in the chamber by replacing it with sterile compressed air at a pressure that is equal to, or higher than, the sterilization pressure. Only after this has occurred does the cooling water spray described in Item 3 begin. This method only prevents the load from suffering the pressure stress of the cooling phase, whereas the stress of the sterilization phase is unavoidable. Reference is made to the section on *Counterpressure Methods* below for an explanation of this phenomenon and for the autoclaves that allow to avoid it.

6. **Spontaneous Cooling**—In some particular cases it may be necessary to resort to this cooling method, which is the simplest but also obviously requires a very long time. Clearly, at the end of this cooling the autoclave will be in vacuum, and the longer the cooling the deeper the vacuum.

7. **Ampule Tightness Test with Dye Solution Penetration**—This test generally is performed with an aqueous solution of methylene blue. However, it is also possible to use other dyes. This test is effective only on *open defects* of ampules and is performed as follows:

   a. Vacuum in the chamber to approximately 100 to 150 mbar.
   b. The chamber is filled with the colored solution until the load is completely covered; the ampules must of course be contained in appropriate trays that do not allow them to escape, because they tend to float.
   c. During this filling operation, the chamber vacuum reached in Item 1 is maintained continuously by connecting the vacuum pump to the ceiling of the chamber.
   d. The colored solution is pressurized at 2 to 3 bar and is maintained in this condition for 30 to 60 min or more.
   e. The colored solution is discharged and recovered.
   f. The load is washed several times with spray water.
   g. The load is washed by flooding the chamber.
   h. The washing water is discharged.

There are alternatives to this method, such as electronic spark discharge inspection which detects leakage of liquid from the ampule by a decrease in resistance across electrodes placed across the ampule.

- The vacuum is not maintained continuously while the chamber is being filled with the colored solution.
- The vacuum is produced only after filling the chamber with the colored solution.
- The vacuum is not produced at all.

This test has in any case the following problems:

- It has been demonstrated extensively that with usual values for dye concentration, differential test pressure, and test time, tip holes with a diameter of less than 5 to 10 μm allow very small amounts of colored solution to enter. This prevents detection of the coloring of the ampules during subsequent checking.
- The preparation of sterile colored solution for each test entails very high costs.
- Recovery and reuse of the colored solution entails keeping it in conditions that prevent microbial proliferation (80°C) and subjecting it to sterilizing filtration prior to each test. All these procedures are expensive and complicated. In any case, the solution recovered from each test is contaminated chemically by the broken or defective tested ampules.
- Decolorization/destruction of the solution is very difficult, because methylene blue is very stable; however, good decolorization results have been achieved by using ozone. The use of amber glass ampules makes detection of the dye difficult.

STERILIZING THE AIR INTRODUCED IN THE CHAMBER

In the previous paragraphs we noted that it is often necessary to introduce air in the chamber, especially in post-sterilization phases. This air must be sterile, otherwise it may recontaminate the sterilized load and can, in any case, contaminate the sterile environment if the autoclave is of the two-door type connected to the sterile area.

The air generally is sterilized by filtration using a system that is part of the autoclave. It is thus necessary to

- Provide a filtration cartridge with sterilizing porosity
- Allow *in situ* sterilization of the assembled filtration system with an appropriate sterilization program of the autoclave itself
- Ensure that the filtration system and the line for connecting it to the autoclave maintain their sterility between one production sterilization program of the autoclave and the next
- Allow validation of all of the above described procedures

If one wishes to operate in perfect safety, the filtration system also should be subjected to an integrity test each time it is operated.

Counterpressure Methods

Autoclaves operating with counterpressure are defined as devices able to control, during sterilization, the pressure of the moist sterilizing medium independently of its temperature. Conventional pure saturated steam autoclaves do not belong to this category. The temperature of the pure saturated steam present in the chamber in fact automatically generates a specific pressure that cannot be modified without modifying the temperature as well. If the temperature of the steam is 121°C, its pressure is unavoidably 2.05 bar abs and *vice versa,* assuming no trapped air.

For many kinds of load it is necessary or convenient to use an autoclave operating with counterpressure. To understand this need, let us see what happens in a conventional autoclave during the sterilization of a rigid container partially filled with an aqueous solution and closed tight. For the sake of simplicity, let us assume that the container is filled with pure water.

A glass bottle is filled partially in standard conditions: 20°C and 1.013 bar; the bottle is closed with a rubber stopper and aluminum seal. In the head space there is a total pressure of 1.013 bar, which is actually the sum of two factors: a partial water-vapor pressure which corresponds to the vapor pressure of water at 20°C, ie, 0.025 bar, and a partial air pressure of 0.988 bar.

When the bottle is subjected to the sterilization phase at 121°C, these two factors change as follows:

	Initial Condition		Sterilization Condition
Partial water-vapor pressure	0.025	→	2.050 bar (1)
Partial air pressure	0.988	→	1.330 bar (2)
Total pressure in head space	1.013	→	3.380 bar abs

Value 1, 2.050 bar, is obviously the pressure of water vapor at 121°C and *corresponds to the pressure that occurs in the autoclave chamber.* Value 2, 1.330 bar, is a theoretical value that is calculated by applying the law of perfect gases to air:

$$0.988 \times \frac{121 + 273}{20 + 273} = 1.330$$

Therefore, the total pressure of 3.380 bar abs is also a theoretical value.

There are some reports that demonstrate that the *practical* value is slightly higher than the theoretical one and largely depends on the ratio between the head space and the volume of the filling solution. The practical pressure of the head volume is, on average, higher at 121°C by approximately 1.40 bar, with respect to the pressure in the chamber. This is caused by two mechanisms:

The thermal expansion of water is significantly greater than that of glass and increases very rapidly as the temperature rises. The specific volumes of water at the temperatures we are interested in are in fact

Temperature °C	Specific Volume mL/g
0	1.0002
4	1.0000 (maximum density)
20	1.0017
120	1.0606

In passing from 20° to 121°C, water increases its volume by approximately 6% according to the following ratio:

$$\frac{1.0606}{1.0017} = 1.058$$

This fact must be considered carefully by those who tend to reduce or eliminate the head space in containers and then are surprised to find that such containers explode or warp during sterilization. Solutions (especially if filtered under gas pressure) contain considerable amounts of dissolved gases that leave the liquid phase as the temperature rises.

The overpressure of approximately 1.40 bar that occurs in the bottle naturally generates a force of approximately 1.4 kg per cm^2 of internal surface of the bottle. A rubber stopper with a diameter of 24 mm is subjected to an expulsion force of approximately 6.3 kg.

These conditions therefore prevent or advise against the use of a pure saturated steam autoclave to sterilize solutions contained in a wide variety of containers. For example,

- Large-Volume Parenterals (LVP) in glass containers
- Small-Volume Parenterals (SVP) in glass vials with rubber stopper
- LVP or SVP in plastic containers (flexible, semirigid, or rigid plastic)
- Prefilled syringes
- Jars or similar containers with press-on or screw on closures
- Blisters containing various materials, such as disposable contact lenses

Two counterpressure methods currently in use are

- Superheated water spray method (water cascade process)
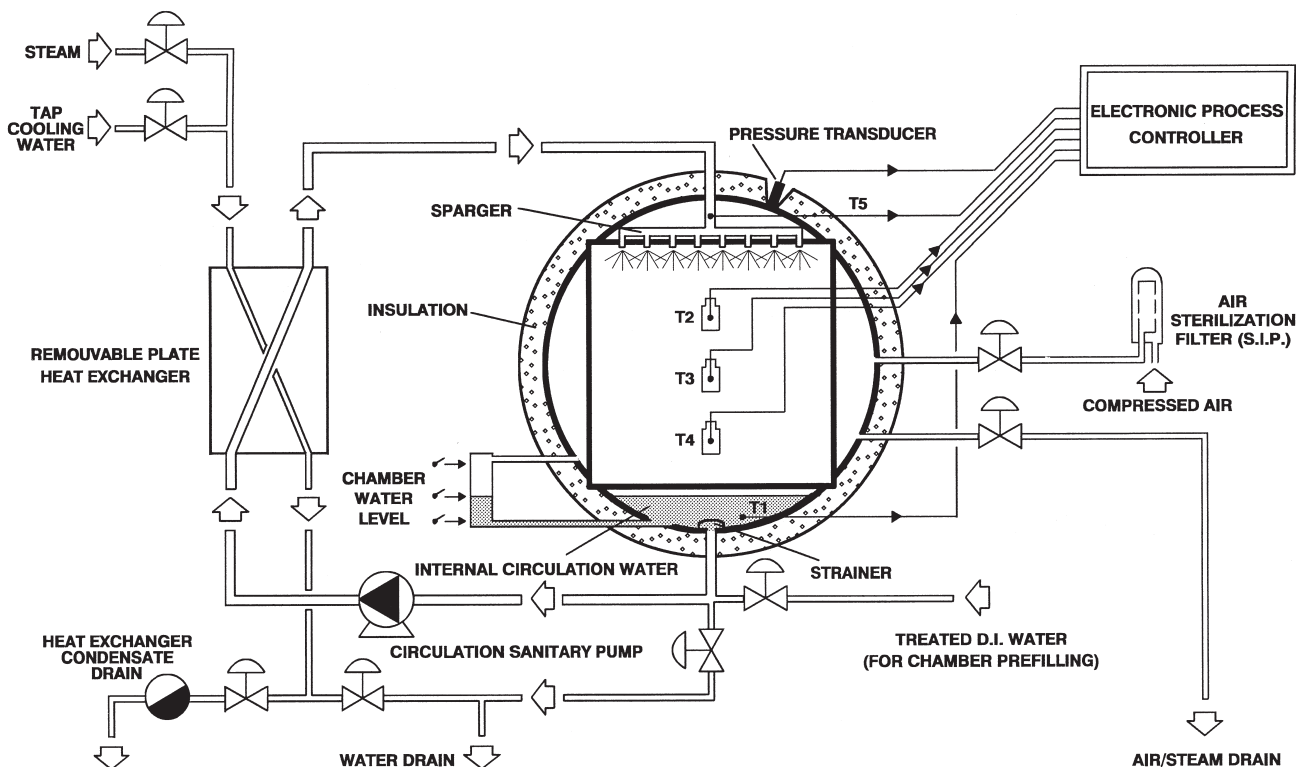- Air overstream method (steam plus air method)

Figure 40-5. Superheated water-spray autoclave: simplified piping and instrumentation diagram (courtesy, Fedegari Autoclavi).

SUPERHEATED WATER SPRAY AUTOCLAVES

A typical functional diagram of this autoclave is shown in the Figure 40-5. Obviously, different solutions are also possible which, however, do not change the essence of the method. The chamber generally has a circular cross-section (with quadrangular door(s) inscribed in the circumference) and has a single wall.

At the beginning of the program, after the goods have been loaded, the lower circular sector is filled with purified water. The air contained in the chamber is *not* removed. The water, drawn by a sanitary-type pump, circulates in a heat exchanger (plate or other sanitary type), which is *indirectly* heated in countercurrent with industrial steam. The water returns then into the upper part of the chamber and is distributed to the load by a system of solid-cone spray nozzles. The uniform redistribution of the water on the lower layers of the load is ensured by appropriate perforated racks that support the load. Side spray bars sometimes are used, even if their actual usefulness is not demonstrated.

The heating of the circulation water, and therefore of the load, is gradual but quite fast; for example, the temperature of 121°C is reached in approximately 20 to 30 min *inside* 500-mL containers, mainly dependent on the solution and the material and shape of the containers.

The sterilization phase lasts 15 to 20 min, and temperature uniformity (in time and space) is excellent: it is well within the quite narrow limits required by FDA for LVP sterilization, ±0.5°C. This allows very small F_0 dispersions, and therefore minimum sterilization times.

The cooling phase is performed while the circulation water, now sterile, continues to circulate. However, cold tap water now flows in the plates of the exchanger, where steam was flowing earlier. In less than 15 min, the temperature *inside* the 500-mL containers drops to approximately 70°C, which is also the ideal temperature for obtaining a rapid and spontaneous drying of the load removed from the autoclave.

During all the phases of the process, an appropriate sterile air counterpressure is maintained inside the chamber to counterbalance the overpressure in the bottles. There are various methods for controlling this counterpressure in each phase. With computerized management, it is even possible to generate a total pressure (steam plus air) inside the chamber that is correlated, in each phase, to the average of the internal temperatures of two or more *witness* containers.

The load suffers no thermal or pressure shock and the differential pressure between containers and chamber can be eliminated or maintained in a direction convenient, in each phase, for the particular type of load. Even highly deformable products (semirigid plastic containers or plastic–aluminum blisters) or products that are particularly sensitive to differential pressures (eg, prefilled syringes) can be treated (from 60° to 127°C) without problems.

The autoclaves are obviously highly specialized machines, and as such they have some limitations in application:

- It is illogical to attempt to dry the load inside the autoclave by putting the chamber in vacuum or by circulating warm air.
- In the case of materials with concavities directed upward, these concavities will be filled with water at the end of the program: the most obvious solution is to load these materials upside down.
- When PVC bags are sterilized, the phenomenon of *blushing*—the whitening of the PVC due to water absorption—usually occurs. The intensity of this phenomenon and the time required for its disappearance depend on the type of PVC and of plasticizer employed. Blushing does not occur with rigid or semirigid plastic or with polylaminate plastics; it also is reduced considerably with PVC containing special plasticizers.

AIR OVER STEAM AUTOCLAVES

A typical functional diagram of this type of autoclave is shown in Figure 40-6. Alternatives are also possible in this case. The most important one is the use of horizontal fans placed on a side of the chamber. As in the previous case, the chamber has a cir-
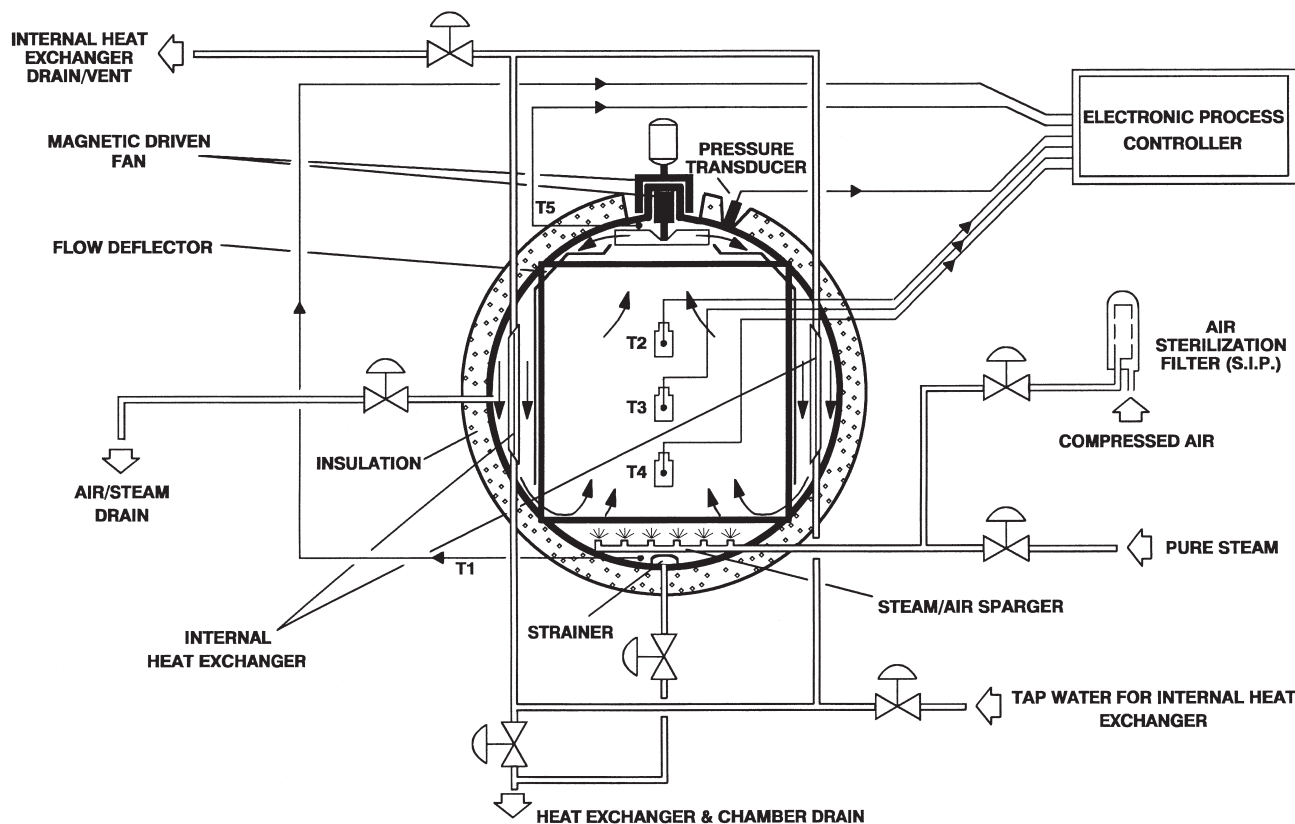
Figure 40-6. Air over steam autoclave: simplified piping and instrumentation diagram (courtesy, Fedegari Autoclavi).

cular cross-section (with a quadrangular door or doors inscribed in the circumference) and has a single wall.

There are two reasons for choosing a circular cross-section for autoclaves operating in counterpressure.

1. These autoclaves operate at significantly higher pressures than conventional pure saturated steam autoclaves, and generally are not put in vacuum. It is well known that a cylindrical structure withstands internal pressure much better than a quadrangular one.
2. The circular sectors of the chamber that are not occupied by the load are used to place elements required for the operation of these autoclaves.

The air is *not* removed initially from the chamber. The steam enters directly into the chamber through a sparger located in the chamber's lower portion. The partial air pressure of the mixture can be adjusted during the entire process, similarly to what occurs for the previously described superheated water spray autoclaves.

The fan(s) placed against the ceiling of the chamber and the flow deflectors have the purpose of homogenizing the steam plus air mixture that forms inside the chamber. The task of these fans is very important and demanding. In fact, for equal pressure and temperature conditions, the air is approximately 1.6 times denser than the steam (one only has to consider their respective molecular weights) and would tend to stratify on the bottom, producing intolerable temperature gradients.

The cooling phase consists of feeding air into the chamber (to condense and replace all the steam that is present) while maintaining the same sterilization pressure or possibly increasing it. Cold tap water then is fed into the heat exchangers, which are constituted by batteries of hollow plates located in the two circular sectors of the sides of the chamber (only one plate is shown in the diagram for the sake of simplicity). A tube heat ex-

changer can be used as an alternative. The load is thus cooled while constantly maintaining a controlled pressure inside the chamber.

However, this cooling comprises two solid-gas heat exchanges (plates → air; air → load) that, as is known, have a very poor efficiency. An attempt is made to improve this exchange by increasing the pressure of the air in the chamber (within the limits allowed by the product and the autoclave) so as to increase its density and therefore its heat-exchange capacity. The fans obviously continue to run during the cooling phase. Despite these refinements, the cooling phase is definitely longer than the same phase in superheated water spray autoclaves.

A critical mechanical aspect of these autoclaves is the tightness of the fan shaft. This aspect can be solved completely by using magnetic-drive fans.

With steam plus air mixture autoclaves, the blushing of PVC bags is less intense than with water spray autoclaves and generally affects essentially the regions where the bag rests on the supporting racks.

Table 40-1 compares the characteristics of the two kinds of counter pressure autoclaves.

Dry Heat Treatments

STERILIZATION AND DEPYROGENATION

Dry heat treatments have two targets: microorganisms and their by-products. The aim of sterilization is to destroy the ability of microorganisms to survive and multiply. Depyrogenation seeks to destroy the chemical activity of the by-products: pyrogens or endotoxins (these terms do not mean exactly the same thing, but we will consider them to be synonymous for the sake of simplicity).

Table 40-1. Counter Pressure Autoclave Comparison

CRITICAL COMPARISON	WATER SPRAY (WS) AUTOCLAVES	AIR OVER STEAM (AS) AUTOCLAVES
Temperature uniformity in time	Very good easily in $\pm 0.5°$ limits	Very good easily in $\pm 0.5°$ limits
Temperature uniformity in space	Very good requested by FDA for LVP	Very good requested by FDA for LVP
Total pressure uniformity in time	Very good	Very good
Counterpressure management flexibility	Excellent	Excellent
Consumption of high microbiological quality water	Yes, modest, for initial filling	No
Consumption of tap water for cooling	Yes, acceptable	Yes, approx. 3 times higher than WAS
Consumption of compressed air	Yes, acceptable	Yes, acceptablke
Consumption of industrial steam	Yes, acceptable	No
Consumption of ultraclean steam	No	Yes, acceptable
Condensate recovery	Possible and easy	Not possible
Cooling water recovery	Possible, recovered water is initially very hot	Possible, recovered water is initially very hot
Autoclave price	Acceptable	Approx. 1.1 times higher than WS
Total process duration	Short	Approx. 1.3 times higher than WS
Autoclave productivity/price	High	Approx. 70% of WS
Operation principle	Very simple and straightforward	More complicated than WS
Mechanical construction	Simple	More complicated than WS
Qualification/validation	Normal	Normal
Operating flexibility according to type of load	Suitable for any kind of container with the following remarks: • Upward concavities collect water • Product is unloaded wet • PVC bags can produce blushing phenomena	Suitable for any kind of container: • Upward concavities collect condensate only • Other kinds of container can be unloaded slightly damp • Blushing phenomena of PVC bags are limited
Possibility of combination with pure saturated steam processes	Strongly discouraged: It is complex and expensive and complicates validation	Very frequent, but moderately expensive

Both processes consist of an oxidation that is almost a combustion. However, the temperatures required to achieve depyrogenation are distinctly higher than those needed to obtain sterilization. We can summarize the situation as follows:

- If an effective dry heat depyrogenation is performed, sterilization generally is achieved *as well*.
- Effective dry heat sterilization can be performed even *without* achieving depyrogenation.
- If moist heat sterilization is performed, in normal operating conditions depyrogenation is *not* achieved.

The kinetics of dry heat treatments are not substantially different from those of moist heat sterilization. The values of the algorithms F_T and F_H (analogous to F_0) and those of the parameters D and z, however, are different not only from those of moist heat sterilization but also from each other. Furthermore, the two dry heat treatments are verified biologically with different biochallenges. Accordingly, the two dry heat treatments require different validation approaches.

The materials subjected to dry heat treatments naturally must be heat-stable: the most common are glass containers for parenterals. Elastomeric compounds generally are unable to tolerate these treatments.

The literature generally mentions the following operating conditions:

Sterilization:	160°C—120 to 180 min
	170°C—90 to 120 min
	180°C—45 to 60 min
Depyrogenation:	230°C—60 to 90 min
	250°C—30 to 60 min

However, the current trend is toward using treatments at higher temperatures than those listed.

The sections that follow describe the most common types of equipment used to perform the above processes. If the load (bottles/vials/ampules made of glass or other materials) is wet when it is introduced, a large part of the energy required by the process is used initially to evaporate the water that wets the load, and the process accordingly takes more time. The equipment uses large amounts of air, which generally is recirculated partially and must be filtered in HEPA filters to have, in the critical regions of the equipment, the Class 100 environment. This is relatively easy to achieve in the sterilization phases (or regions) in which the *thermal situation* of the filters is stable. It is much less easy to achieve in the heating/cooling phases (or regions), because the changes in temperature entail expansions/contractions of the filters, with consequent release of particles.

DRY HEAT BATCH STERILIZERS

The forced-convection batch sterilizer is a type of dry heat unit widely used in the industry. It uses the principle of convective heat transfer to heat the load. Figure 40-7 is a schematic diagram of a modern unit. It shows a two-door sterilizer in which the unloading door leads to the sterile area. The two doors are, of course, parallel to the plane of the drawing and are hinged vertically.

The pressure inside the chamber must be controlled continuously so that it is slightly higher than the pressure in the loading area (nonsterile) and slightly lower than the pressure in the unloading area (sterile).

The unit is made entirely of stainless steel; particular care must be taken in selecting the insulating materials and in the methods for applying them. It is important also to avoid the forming of so-called *thermal bridges;* these allow dissipation, and thus excessive external temperatures of the sterilizer and *cold spots* in the chamber.

The main features shown in the sketch are:

1. Air-circulation fan
2. Water-cooled battery (for the cooling phase)
3. Circulation HEPA filters
4. Launch/recovery bulkheads
5. Trolley and load
6. Discharge duct
7. HEPA filter on the discharge duct to prevent back-flow contamination
8. Variable-speed fan for chamber pressurization (proportionally controlled)
9. Prefilter and HEPA filter on the chamber pressurization loop
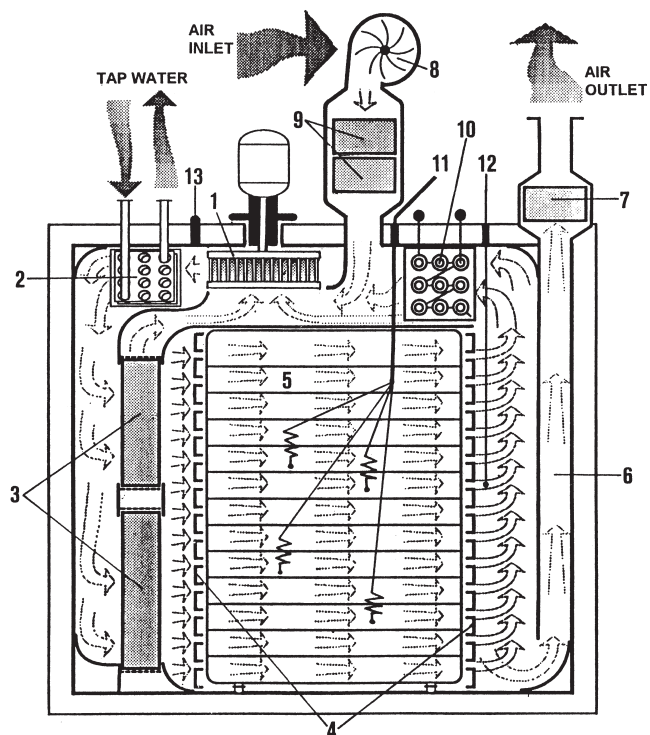10. Electric heater (proportionally controlled)

**Figure 40-7.** Dry heat batch sterilizer: simplified diagram (courtesy, Fedegari Autoclavi).

11. Four flexible Pt100 4-wire RTDs
12. Main control Pt100 4-wire RTD
13. Pressure transducer

DRY HEAT TUNNELS

The drying, sterilizing/depyrogenation, and cooling tunnel is the only continuous sterilizing apparatus widely used in the pharmaceutical industry (apart from filters). It basically consists of a horizontally rotating transport belt made of a stainless-steel mesh (some devices must be provided to confine the product on the transport belt without particulate generating friction), installed in a thermally insulated *tunnel* that directly connects an upstream cleaning machine to the downstream sterile area or to *isolated* devices.

Inside the tunnel, the product (most frequently glass vials) is dried; heat-treated either by radiant heat or, as more usual today, by hot air; and finally cooled. In both cases the internal part of the tunnel must be pressurized dynamically by ventilation at an intermediate pressure level between the downstream system and the loading room. From a process point of view, higher temperature and shorter exposure time are used than in batch sterilizers. During the last 10 years the practice has changed from 20 min at 280°C to 3 or 4 min at 300°C or more. Because a minimum safety margin is required for the duration of exposure, and glass of most types becomes more difficult to handle above 320°C and more fragile after such a treatment, it is likely that the trend toward higher temperature values has reached its practical limit.

In infrared (IR) radiant heat tunnels, heat is supplied by resistance-in-glass heaters located above and below the transport belt; prefiltered and HEPA-filtered air is fed into the cooling zone mainly for pressurizing and cooling. This air, a countercurrent slowly flowing through the entire tunnel, has also an important drying and preheating effect of the load in the infeed zone. Figure 40-8 schematically represents an IR
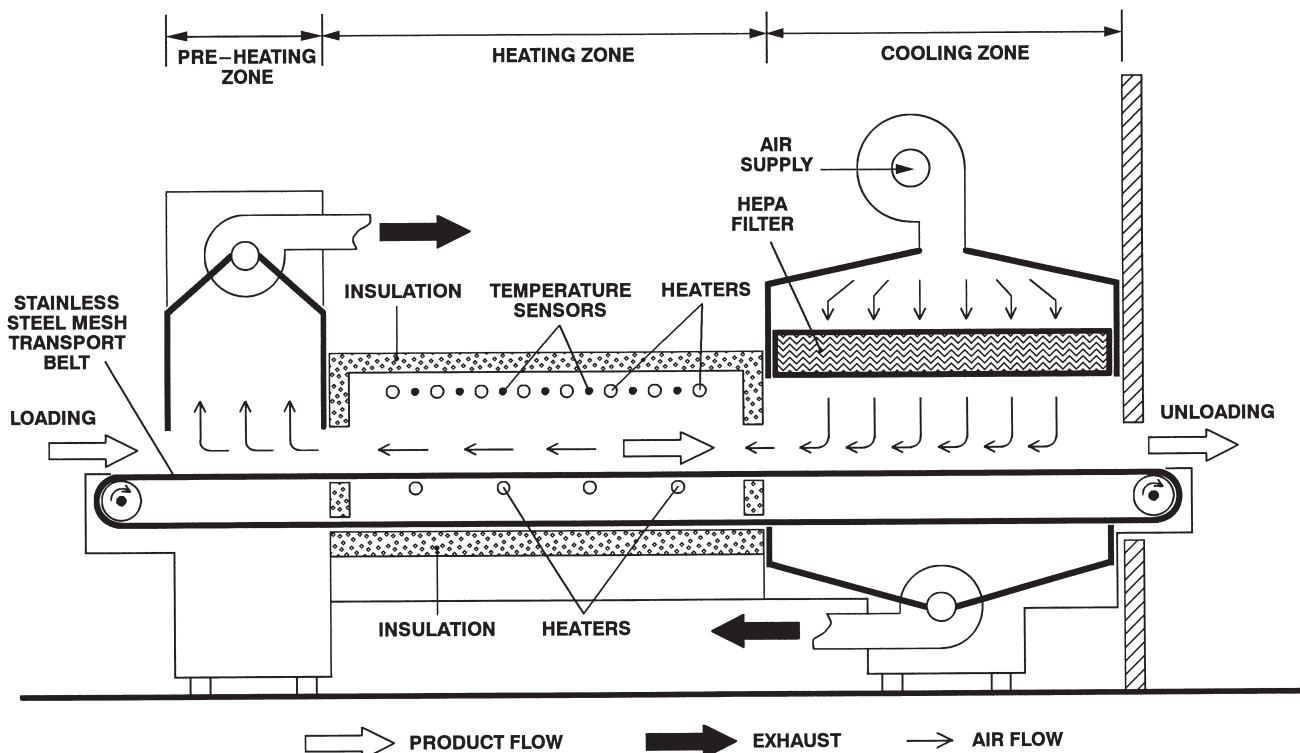
Figure 40-8. Dry heat tunnel: simplified diagram (courtesy, Fedegari Autoclavi).

tunnel: even if this type of apparatus is no longer widely used, the basic concepts have not been modified in the hot-air laminar flow tunnel, but airflow patterns are a little more complex.

Hot-air laminar flow (LF) tunnels do not radiate heat directly to the product, but rather heating is provided by circulation of hot filtered air forced onto the product. A circulation fan withdraws the air; it leaves the product through heating bars below the transport belt and is fed again to the inside of the tunnel through HEPA filters suitable for operating at high temperature. Airtightness of the coupling of HEPA filters with tunnel framework is of utmost importance from the point of view of particulate contamination. It must cope with the strong thermal expansion of different materials. Some makeup air is required in the heating zone, and the total number of installed fans may be as high as five or even six if an additional extraction below tunnel outfeed is required in case of high pressure in the sterile room.

Despite the complexity of its airflow, the LF tunnel has the main advantage of quicker heating and consequent shorter process time. This results in reduced size compared with the IR tunnel, because the belt speed cannot be reduced below a certain value. As the name itself declares, the air speed in the LF tunnel is kept around 0.5 m/sec (1.5 ft/sec), aiming to avoid particulate contamination.

The comparison between continuous tunnel and batch oven is favorable to the continuous tunnel from the point of view of handling the product. No batch work is needed after the unpacking of the components and loading of them into the cleaning machine until the final removal of the packaged product from the line after the filling and the following operations. This can be very important in the case of large-scale production.

The batch oven provides a much easier isolation of the sterile area. In the case of continuous tunnel, there must be a steady flow of air through the open connection from the sterile area to the tunnel. The pressure difference between the two systems must be such that the sterile area always is kept at a higher pressure level than the tunnel. Too big a difference would result in an excessive escape of air to the tunnel, both reducing the pressure in the sterile area and disturbing the laminar airflow and the temperature profile inside the tunnel. Experience has proved that these problems can be solved satisfactorily only if the design of the air-conditioning system of the sterile area is developed from the very beginning, keeping in mind the foreseen installation of a specified tunnel. Baffle systems also aid in maintaining pressure differentials between the aseptic-processing area and the sterilizing tunnel.

Chemical "Cold" Sterilization

Many products do not tolerate the sterilization conditions of moist-heat or dry heat processes. In such cases it is possible to resort to cold or at least low-temperature sterilization methods performed with chemical means, by gases or vapors. The continuously increasing use of plastic disposable products or components for medical treatments has been made possible by the development of reliable cold sterilization processes.

A variety of gases and vapors have shown germicidal properties: chlorine dioxide, ethylene oxide, propylene oxide, formaldehyde, betapropiolactone, ozone, hydrogen peroxide, peracetic acid, etc. Ethylene oxide (EtO) is currently in widespread use for medical product sterilization. However, EtO has been shown to have detrimental effects on the environment; thus, other agents are being developed on a commercial scale, with the intent to reduce the use of EtO. Vaporized hydrogen peroxide and hydrogen peroxide/steam mixtures are being used to sterilize a variety of materials and work surfaces. Chlorine dioxide recently has become available for these applications.

ETHYLENE OXIDE

The sterilizing action of EtO is based on an alkylation reaction: it is, accordingly, a truly chemical action rather than a physical one. This chemical reaction must be activated by the presence of water vapor (approximately 60% of RH or relative humidity) and is increased by temperature and EtO concentration.

The process temperature is limited by the characteristics of the product. Generally, it is between 40° and 60°C, but it must be remembered that the reaction rate increases by approximately 2.5 times for each 10°C increase in temperature. The normally used EtO concentrations range between 400 and 1200 mg/L. It has in fact been demonstrated that beyond 1200 mg/L the consequent increase in the reaction rate is no longer economically convenient.

The EtO must make *direct* contact with the microorganism for the microbe to be inactivated. Any packagings that contain the object to be sterilized must therefore be permeable to air, EtO, and any dilution gases (as discussed later). Generally, it is not possible to use EtO to sterilize liquids, solutions, or emulsions. Powders, too, are difficult to treat unless microbial contamination is only on the outside of the granules.

Fortunately, EtO, air, and dilution gases easily penetrate most of the plastic and paper barriers used for the packaging of medical products. However, the good penetrating properties of EtO are also a disadvantage, because large amounts of it are absorbed by plastic or rubber materials. Products sterilized on an industrial scale using EtO normally require about 14 days of quarantine to spontaneously eliminate absorbed EtO residuals. This time can be reduced by using forced desorption methods. Sterilized goods must be monitored for toxic EtO residual, ethylene glycol, and ethylene chlorhydrin breakdown products of EtO.

EtO in standard room conditions is a vapor (indeed, its boiling point is about 11°C at atmospheric pressure). It is colorless, heavier than air, and has an ether-like odor. Its formula is

$$CH_2\!\!-\!\!\!-\!\!CH_2$$
$$\lfloor\!\!-\!\!O\!\!-\!\!\rfloor$$
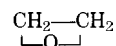
The presence of the oxygen bridge, which can be opened easily, explains its reactivity and its sterilizing action, as well as its tendency to polymerize.

Unfortunately, EtO has several drawbacks: it is toxic, carcinogenic, teratogenic, inflammable, and explosive when mixed with more than 3% air by volume. These characteristics make the use of EtO highly controversial, and many countries have issued regulations or requirements for its use as a sterilizing agent.

EtO often is used in a mixture with dilution gases, with weight ratios of 85 to 90% of diluent. The diluent gas most often used is CO_2; use of Freon is shrinking, due to the well-known international restrictions to its use; N_2 is beginning to be used quite often. These mixtures are considered nonflammable and nonexplosive, and many countries consider them mandatory for use in industrial autoclaves. These countries allow the use of pure EtO in small, individual, single-use cylinders only for small autoclaves (100 to 200 L). If a 10 to 12% mixture of EtO in CO_2 is used to obtain an acceptable EtO concentration (at least 500 mg/L), it is necessary to work at 3 to 4 absolute bar. Accordingly, one must use an autoclave that can withstand relatively high pressures; these autoclaves are expensive, and the duration of the sterilization process is rather long because of EtO concentration is rather low.

Other countries accept the use (including industrial use) of pure EtO or of inflammable/explosive mixtures with a low percentage of dilution gas (the presence of the dilution gas generally is ascribed to a reduction in the tendency of pure EtO to polymerize). In such circumstances one can operate at less than atmospheric pressures and still reach high EtO concentrations that shorten the sterilization time. Thus, it is not necessary to use true autoclaves, but merely sterilizers capable of tolerating the very hard vacuum required for the initial elimination of the air from the chamber and from the load and for the final ex-
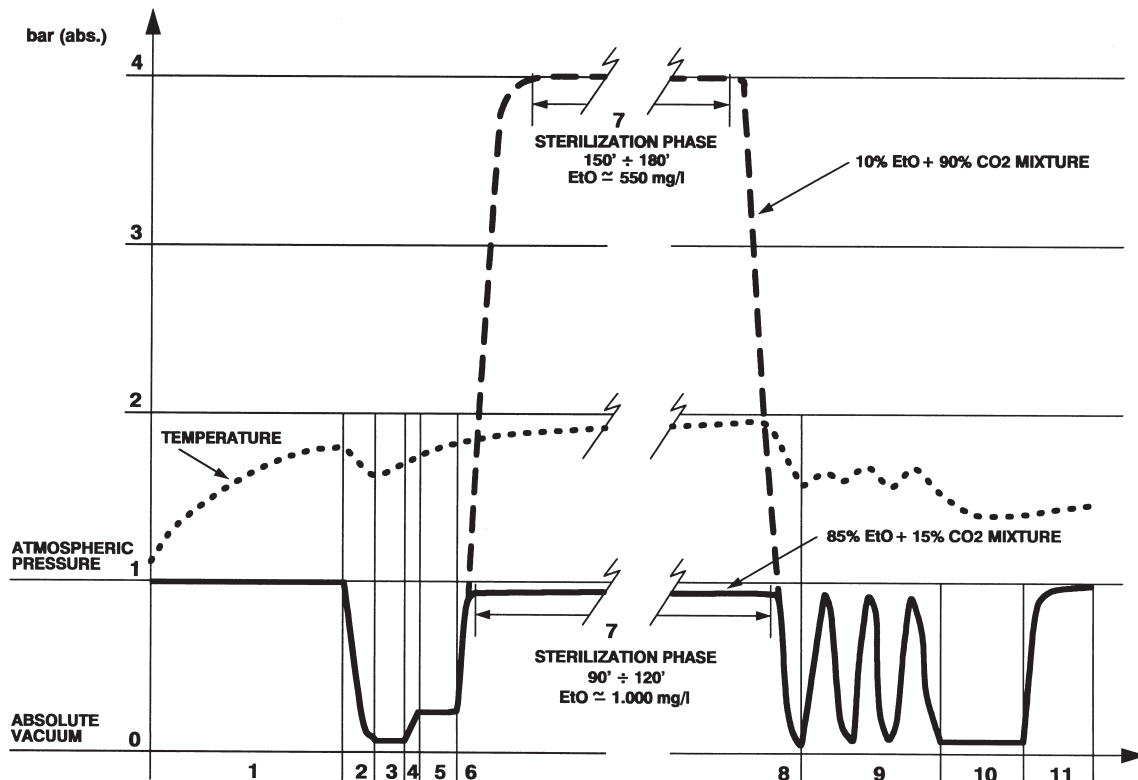
Figure 40-9. EtO sterilization pressure–time diagram: overpressure and subatmospheric pressure (courtesy, Fedegari Autoclavi).

traction of the EtO. Obviously, in these circumstances the use of plants constructed with explosion-proof criteria cannot be avoided.

The P/T/t diagrams of EtO sterilization are therefore different, depending on whether one or the other of the above described principles is used. A typical diagram of an overpressure sterilization with a mixture using 10% EtO and 90% CO_2 is shown in Figure 40-9. These are the steps:

1. Load and/or chamber heating
2. Vacuum
3. Vacuum hold for leak test
4. Humidification by steam injection
5. Penetration of humidity in the load
6. Loading of EtO mixture
7. Sterilization
8. EtO mixture evacuation
9. Air/vacuum pulses
10. Vacuum hold
11. Vacuum breaking

A typical diagram of a subatmospheric sterilization with a mixture using 85% EtO and 15% CO_2 is shown equally in Figure 40-9. One can see clearly that the phases are substantially the same as in Figure 40-8; the changes are the sterilization pressure, the EtO concentration, and therefore the duration of the sterilization phase.

In performing industrial sterilizations, which accordingly involve large loads, the load is heated and humidified before placing it in the sterilizer, in adequately conditioned rooms. Thus, the heating/humidification phases described above in the diagrams of Figures 40-8 and 40-9 are reduced drastically.

The layout of an industrial EtO sterilization plant is shown in Figure 40-10. This unit contains:

• The EtO or EtO-mixture cylinders
• The automatic devices that connect/disconnect the various cylinders to and from the sterilizer; disconnection of a cylinder (especially for mixture cylinders) often is controlled by its weight reduction, which must accordingly be checked individually
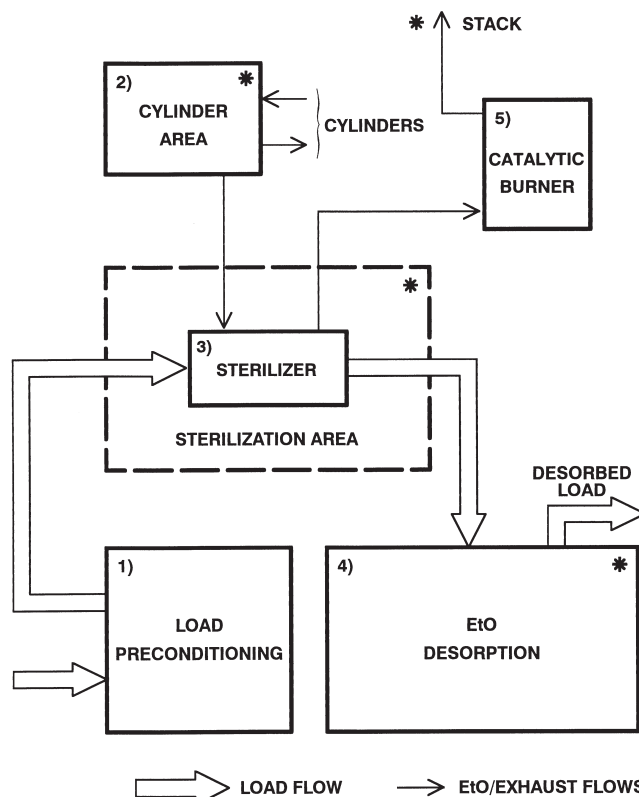

Figure 40-10. Flow diagram of an EtO sterilization industrial plant (courtesy, Fedegari Autoclavi).

- The heat exchanger that must provide the vaporization calories to the liquid EtO mixture
- The pressure reduction unit that brings the liquid EtO mixture to the vapor state
- Any cylinders of N_2, which is used in the most advanced plants to *wash,* after each process, the pipes that have carried EtO

The EtO that is produced in the desorption chamber is at a very low concentration and it is generally too expensive to eliminate it with a catalytic burner. It is preferred to absorb it on activated-charcoal columns through which the air of the desorption chamber is recirculated.

Obviously, the EtO discharged by the sterilizer (and possibly the EtO arriving from the desorption chamber) must not be discharged into the atmosphere. Catalytic burners generally are used today: they convert the EtO into $CO_2 + H_2O$. These burners must be highly efficient, and their efficiency must be checked systematically, because the laws enforced in the various countries are generally very strict as to the limits of residual EtO. The asterisks (*) in Figure 40-10 indicate points where continuous monitoring of EtO concentration must be provided.

EtO sterilizers generally are made of stainless steel, although there also are machines made of carbon steel coated with epoxy paints. The chamber generally is jacketed in order to circulate thermostat-controlled warm water to maintain the sterilization temperature. Use of water vapor for the same purpose is fading because of the difficulties in using this method when the temperature must be kept below 100°C.

Process sensors are more numerous than in heat sterilizers, because there are essentially four sterilization parameters:

- EtO concentration
- Temperature
- Humidity or relative humidity
- Time

The EtO concentration generally is monitored by the pressure rise that occurs in the chamber when the EtO mixture is introduced; a pressure transducer is therefore used as a sensor.

Many guidelines require, in addition to the pressure rise, a second monitoring method that can be chosen among the following:

1. Weight difference of gas cylinder
2. Volume of gas delivered
3. Sampling from the sterilizer chamber and analysis

When EtO mixtures are used, Methods 1 and 2, like the pressure rise method, assume confidence in the concentration of EtO that is present in the mixture and that reaches the sterilizer. Method 3 is certainly more reliable, but it also is more difficult to apply. Infrared spectrometry or gas chromatography methods generally are used for analysis; these methods can be continuous and allow the simultaneous determination of the relative humidity.

Temperature generally is monitored by temperature sensors located in the chamber; these may be placed inside the load as well. The relative humidity generally is monitored on the basis of the temperature and of the pressure rise of the steam injection of the humidification phase. This method obviously is not very reliable, and many guidelines recommend also using a sensor that can determine the relative humidity. Unfortunately, sensors of this kind are generally "poisoned" by the EtO and become unreliable after a few cycles. The solution often used is to keep the sensor inside the chamber during the humidification phases, removing it before introducing the EtO in the chamber.

Finally, it is evident that if the load is preconditioned, the preconditioning temperature/humidity/time also must be monitored.

HYDROGEN PEROXIDE

Hydrogen peroxide (HP), chemically H_2O_2, is normally a liquid at room temperature. However, it can be vaporized and the resultant gas is an effective sterilant for certain packaged materials and for equipment and enclosures used in processing sterile materials. The most frequent and successful use of HP as a sterilant is for *isolators* (also known as barriers, locally controlled environments, etc). These units are very sophisticated versions of their ancestors, the *glove boxes* used to isolate processes in the past.

Isolators now are used widely for sterility testing, transporting sterilized goods from moist and dry heat units to sterile areas or processing isolators, and processing of supplies. HP also is being used to sterilize more sophisticated processing equipment, such as freeze dryers and filling lines, and even may be used to sterilized small clean rooms. High humidity can inhibit the effectiveness of vaporized HP and must therefore be controlled during the exposure of the gas. Figure 40-11 represents a typical vaporized HP cycle.

Although HP is broken down readily to water and oxygen, the effluent gas can represent a safety hazard at higher levels. Just as with EtO, catalytic converters are used to ensure that all materials are rendered safe before they are released to the atmosphere. Figure 40-12 represents a typical installation using vaporized HP to sterilize a freeze dryer and condenser system. VHP DV1000 is a model manufactured by Am Sterilizer/Finn Aqua, which holds many of the patents on the use of this technology.

Systems for larger applications may require fans to aid in uniformly distributing the vaporized HP. In addition, auxiliary air systems may be added to reduce the time required to dehumidify at the beginning of the cycle and to aerate the load at the end of the cycle. Figure 40-13 shows a transfer isolator connected to a sterilizer and a vaporized HP generator. This particular unit also has a protective half-suit to allow full access to the large internal area. These units allow the unloading of the sterilizer directly into a sterilized isolator. The isolator excludes direct human intervention, which greatly reduces the potential for microbial contamination.

A typical freeze-dryer sterilization involves several vacuum *pulses* during which the temperature is brought to 40° to 60°C and the humidity is reduced (dry phase). A vacuum hold cycle is run to check for leaks and the temperature is reduced to about 25°C for the sterilization cycle. The sterilant is introduced and is monitored and controlled by weight using an electronic balance. Filtered air is pulsed with sterilant to push the vapors into any deadlegs and to compress the vapors, thus increasing the concentration. Finally, the vacuum is pulsed again to aerate the chamber, and the residual vapor is verified to be below acceptable levels before proceeding to the processing cycle.[11]

HYDROGEN PEROXIDE PLUS STEAM

For certain applications, one can combine moist heat and hydrogen peroxide methods. The combination can produce some effects that may be more desirable than either of the techniques run separately. Cycles can be as effective in shorter times and may improve the removal of residual peroxide. The system must be able to withstand exposure to steam at atmospheric pressure. The air-handling equipment can be moved outside the processing area, which simplifies the system and minimizes any mechanically generated particles, because the air, steam, and peroxide are introduced through the same type of HEPA filters used for laminar-flow hoods.[12,13]

The process area is raised to about 80°C by introducing dry heated air through the HEPA filters. The steam is introduced and surfaces are raised to about 100°C. During the steam cycle, hydrogen peroxide is introduced and is carried with the steam. When the cycle has been completed, the steam and peroxide are stopped and the dry heated air is started again. This aids in removal of residual condensate and helps break down the peroxide to water and oxygen. After sufficient heat has been introduced to dry and remove residuals, cool air is introduced to bring the unit to the desired operating temperature.

Because the hydrogen peroxide is mixed intimately with the steam, temperature can be used to monitor the progression of the cycle. However, the heated portions of the cycle must be
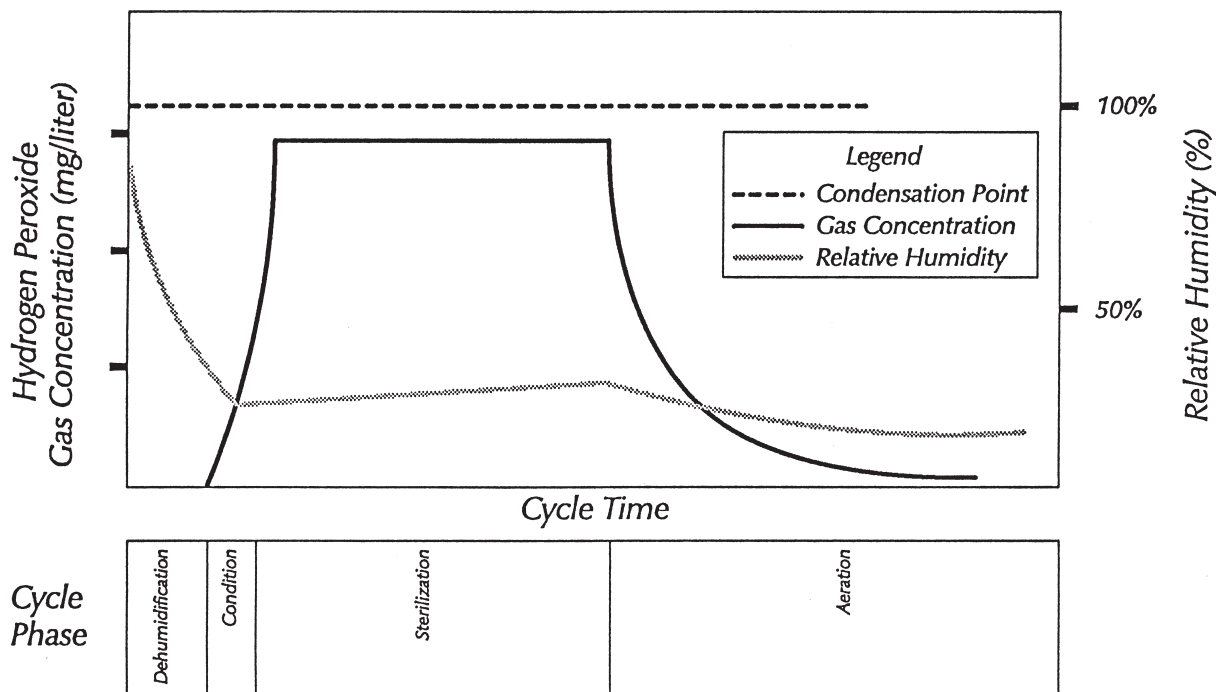
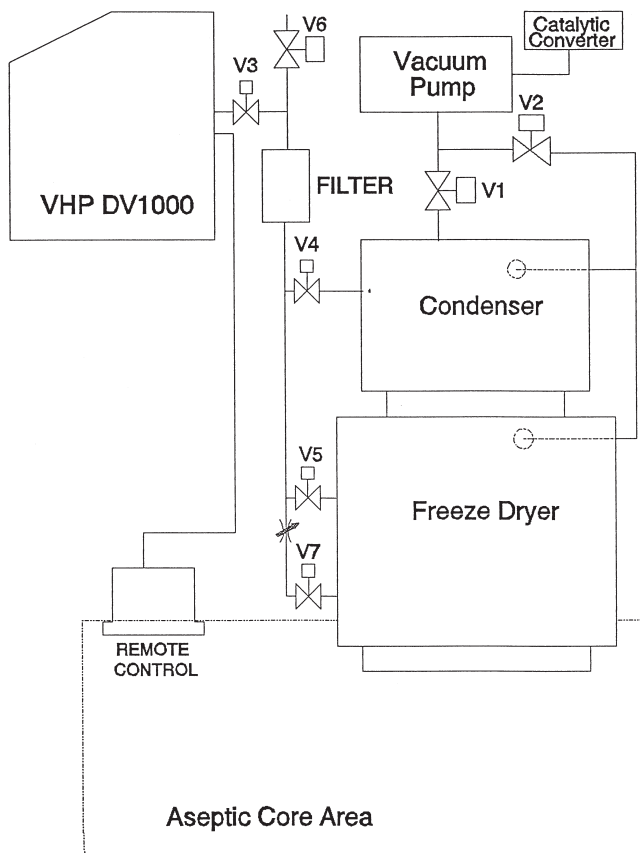**Figure 40-11.** A typical vaporized HP cycle (courtesy, Am Sterilizer/Finn Aqua).

validated using biological indicators and residual peroxide measurements, to ensure their effectiveness in sterilizing and removing residuals to a safe level. Figure 40-14 diagrams a cycle using steam and hydrogen peroxide to sterilize as a filler in an isolator.

Figure 40-15 is included to show the synergistic effects of steam and hydrogen peroxide in some sterilization cycles. The challenge organism was *Bacillus stearothermophilus,* which typically is used to validate steam cycles. It should be noted that the kill rate was not only considerably faster, but was accomplished using atmospheric steam. This means that instead of 121°C the equipment was only subjected to 100°C and was exposed for 15 min less to achieve the same reduction in microorganism count.

Figure 40-12. A typical installation using vaporized HP to sterilize a freeze dryer and condenser (courtesy, Am Sterilizer/Finn Aqua).

Figure 40-13. A transfer isolator connected to a sterilizer and a vaporized HP generator (courtesy, Am Sterilizer/Finn Aqua).
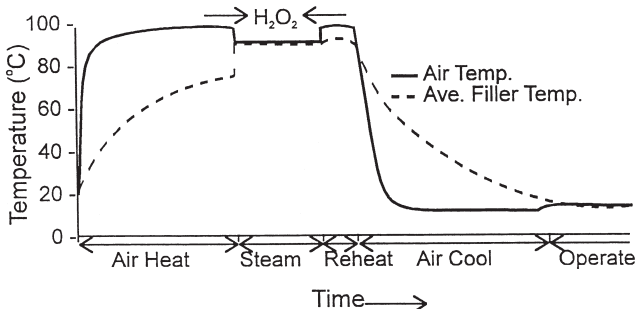
Figure 40-14. A steam/hydrogen peroxide cycle (courtesy, TL Systems and Despatch Industries).

CHLORINE DIOXIDE (CD)

The compound chlorine dioxide (CD) was discovered in 1811. It is a greenish-yellow gas with the common name euchlorine. It is a single electron transfer–oxidizing agent that has a chlorine-like odor. CD has been recognized since the beginning of the 20th century for its disinfecting properties. CD possesses the bactericidal, virucidal, and sporicidal properties of chlorine, but unlike chlorine, does not lead to the formation of trihalomethanes or react with ammonia to form chlorinated organic products (chloramines). These properties have led to the widespread use of CD in the treatment of drinking water. Despite numerous applications for CD in aqueous systems, only recently have the sterilizing properties of gaseous CD been demonstrated.

CD has been shown to have low toxicity in humans and is nonmutagenic and noncarcinogenic; it is not an ozone-depleting chemical. Used at comparatively low concentrations and at subatmospheric pressure, gaseous CD sterilization lacks many of the hazards associated with EtO, and it has been suggested as an attractive potential replacement.[14,15] Gaseous CD does not require expensive damage-limiting construction and is cost-competitive with EtO. Capability for spectrophotometric in-chamber measurement of gas concentration makes the process amenable for the validation of parametric release.

CD gas cannot be compressed and stored in high-pressure cylinders, but is generated upon demand using a column-based solid phase generation system. The chemical reaction used for CD generation is based upon the reaction of solid flaked sodium chlorite with dilute chlorine gas:
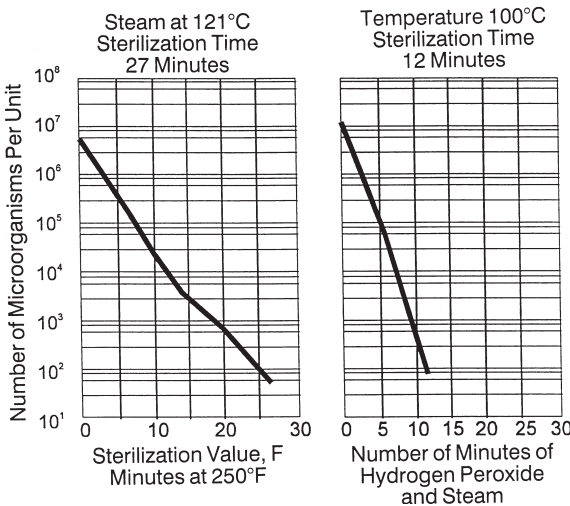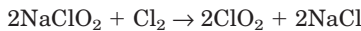
$$2NaClO_2 + Cl_2 \rightarrow 2ClO_2 + 2NaCl$$

Figure 40-15. Comparison of steam under pressure with hydrogen peroxide/atmospheric steam mixture (courtesy, TL Systems and Despatch Industries).

Table 40-2. Effect of CD Gas Concentration on the Rate of Inactivation of 10^6 *B subtilis* Spores on Paper Strips Within a Load of Overwrapped Foil Suture Packages[a]

EXPOSURE PHASE TIME (min)	FRACTION NONSTERILE[b]		
	10 mg/L	20 mg/L	40 mg/L
0	NT	20/20	19/20
15	NT	19/20	1/20
30	20/20	4/20	0/20
60	9/60	0/60	0/20
90	3/20	NT	NT
180	0/20	NT	NT
240	0/20	0/20	NT

[a] The paper spore strips were placed next to the foil suture package and then overwrapped with Tyvek/Mylar. Sterilization exposures were performed at 30 to 32°.
[b] NT = not tested.

A block diagram for a CD gas sterilization system is shown in Figure 40-16. The output of the primary generation column is monitored spectrophotometrically, as is the gas concentration within the chamber. The scrubber system uses a sodium thiosulfate solution to chemically convert the CD to sodium sulfate. The scrubber system is highly efficient; therefore, the effluent released into the atmosphere is mainly process N_2 and air with the CD component reduced to low ppm levels. A typical gaseous CD sterilization process is quite similar to that used with EtO and has these steps:

1. Initial vacuum to remove air from the chamber and load.
2. Moisture conditioning at 70 to 85% relative humidity for 30 to 60 min.
3. CD gas injection: 10 to 30 mg/L.
4. Air or N_2 injection to attain a constant subatmospheric pressure, generally 80 kP_a.
5. CD gas exposure, generally 60 min.
6. Chamber and load aeration by evacuation and air replacement, tailored to load materials and density.

The temperature of the process in a sterilizer application is 30 to 32°C; for isolation systems, it is at ambient temperature.

Feasibility studies on the application of gaseous CD for medical sterilization were performed with over-wrapped foil suture packages.[16] The studies focused on the effect of gas concentration on the rate of inactivation of paper-strip biological indicators (BIs). The results of these studies are shown in Table 40-2. As with other gaseous sterilants, as the CD concentration increases, the time it takes to attain all sterile BIs becomes progressively shorter.

More detailed CD sterilization process development and validation studies were performed using polymethylmethacrylate (PMMA) intraocular lenses as the test system. A diagram of the sterilization process used for these studies is shown in Figure 40-17. The following results were obtained after 30 minutes of gas exposure at 30 mg/L (half cycle):

Packages/Load	Fraction Nonsterile
800	0/8, 0/8
1600	0/16, 0/16
25	0/25, 0/25
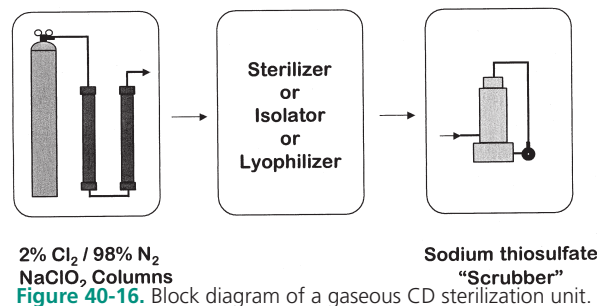
2% Cl$_2$ / 98% N$_2$ **Sodium thiosulfate**
NaClO$_2$ Columns   **"Scrubber"**

Figure 40-16. Block diagram of a gaseous CD sterilization unit.
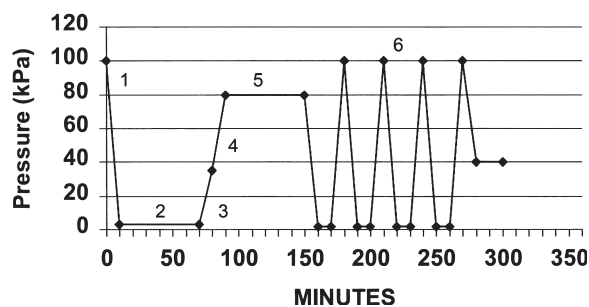
Figure 40-17. Pressure excursion diagram of a typical gaseous CD sterilization process. (1) Initial vacuum; (2) Moisture conditioning; (3) CD gas injection; (4) N_2 or air injection; (5) CD gas exposure phase; (6) aeration by evacuation and air replacement.

As can be seen, all of the *B. subtilis* BTs were sterilized and varying the load size had no discernible effect upon process lethality. CD also has been evaluated for the sterilization of blood oxygenators.[17]

CD also has great potential for the decontamination/sterilization of barrier-isolation systems. Initial studies on the efficacy of gaseous CD for the decontamination/sterilization of a sterility testing isolator used a gas concentration of 10 mg/L. This concentration yielded a relatively rapid process with a complete kill of 10^6 spores in approximately 15 min. The effect of gas concentration upon the observed D_{10} value with *B. subtilis* spores was determined at 10, 20, and 30 mg/L of CD:

mg/L CD	D Value in Seconds
10	45
20	16
30	7

As expected, the D_{10} value decreases with increasing CD concentration. These low D_{10} values yield very rapid decontamination/sterilization processes for barrier-isolation applications.

Very low residuals of CD are observed when examining product and packaging materials from medical devices or isolation technology systems. CD does not appear to have the *solvent-like* quality of EtO. Residual CD is generally less than 10 ppm following a 15-min exposure at 10 mg/L. Rapid aeration also is observed with levels often less than 1 ppm following 15 min of aeration. A typical aeration curve of CD from flexible-wall isolator PVC material is shown in Figure 40-18.

The impact of CD exposure on a number of polymeric materials and metals has been evaluated. Commonly used polymers such as ABS, nylon, PMMA, polyethylene, polypropylene, polystyrene, Teflon, and Viton appear highly compatible. Poly-
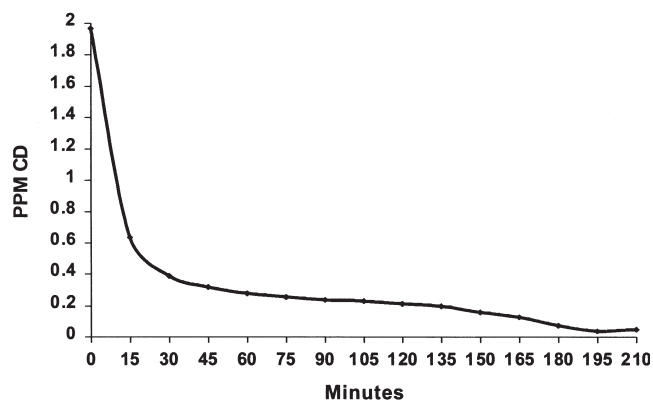
Figure 40-18. Aeration of CD from flexible-wall isolator PVC material; aqueous extraction from treated samples (10 mg/L, 15 min) followed by polarographic measurement of dissolved CD.

carbonates and polyurethanes, depending upon the particular formulation, may exhibit a loss in tensile properties and/or discoloration. Stainless steel is compatible with CD; uncoated copper and aluminum are affected.

OTHER GASES

Formaldehyde (HCHO) sometimes is used for sterilizing certain medical products. It is not in widespread use in the United States but as a gas or in combination with low-pressure steam, it is used in some European hospitals instead of ethylene oxide. Formaldehyde, a toxic chemical and a human carcinogen, is an alkylating agent and destroys microorganisms by alkylation of susceptible cell components.

Filtration

Filtration is the removal of particulate matter from a fluid stream. Sterilizing filtration is a process that removes, but does not destroy, microorganisms. Filtration, one of the oldest methods of sterilization, is the method of choice for solutions that are unstable to other types of sterilizing processes.

Pasteur, Chamberland, Seitz and Berkfeld filters have been used in the past to sterilize pharmaceutical products. These types of filters were composed of various materials such as sintered glass, porcelain, or fibrous materials (ie, asbestos or cellulose). The filtration mechanism of these depth filters is random adsorption or entrapment in the filter matrix. The disadvantages of these filters are low flow rates, difficulty in cleaning, and media migration into the filtrate. Fiber-releasing and asbestos filters now are prohibited by the FDA for the filtration of parenteral products.[18,19]

Over the past 35 years, membrane filters have become the method of choice for the sterilization of heat-labile sterile products. Membrane filters are thin, strong, and homogenous polymeric structures. Microorganisms, present in fluids, are removed by a process of physical sieving and are retained on or near the membrane surface. Membrane filters of 0.1 and 0.22-μm pore size are employed commonly as sterilizing filters.

When solutions are sterilized by filtration, the filters must be validated to ensure that all microorganisms will be removed under known conditions. Filter manufacturers normally validate sterilizing membrane filters using a protocol similar to the one developed by the Health Industry Manufacturers Association (HIMA).[20] In this procedure, *Pseudomonas diminuta* (ATTC 19146) is cultivated in saline lactose broth. Leahy and Sullivan[21] have shown that when *P. diminuta* is cultivated in this medium the cells are discrete and small (approximately 0.3 μm in diameter)—a range recommended for sterilizing filtration with 0.22-μm filters. Each cm² of the filter to be validated is challenged with 10^7 microorganisms at a differential pressure of 30 psig. The entire filtrate is collected and tested for viable microorganisms. The retention efficiency (log reduction value) of the membrane filter may be calculated using the procedure described in the HIMA protocol. Dawson and co-workers[22] have demonstrated that the probability of a nonsterile filtration with a properly validated membrane filter is approximately 10^{-6}. Another aspect in filter validation is adsorption of the product by the filter and extractables from the filter and housing.

Once the performance of the membrane filter has been validated, a nondestructive integrity test that has been correlated to the bacterial challenge test (the bubble point or diffusion test) can be used routinely prior to and after a sterilizing filtration to ensure that the membrane filter is integral.[23,24] Unique to membrane filtration is the condition that beyond a certain challenge level of microorganisms, the filter will clog. For a typical sterilizing filter this level is 10^9 organisms per cm². Initially, membrane filters were available only in disc configuration. Advances in membrane technology have provided filters in stacked-disc, pleated-cartridge and hollow fiber configurations. These advances have provided larger surface areas and higher

Figure 40-19. Stacked-disk membrane filters. This new technology allows filter manufacturers to supply filters with large surface area in relatively small packages (courtesy, Millipore); vaporized HP generator (courtesy, Am Sterilizer/Finn Aqua).

flow-rate capabilities. Figure 40-19 is an example of these larger surface area filters.

Membrane filters are manufactured from a variety of polymers, such as cellulosic esters (MCE), polyvinylidiene fluoride (PVF), and polytetrafluoroethylene (PTFE). The type of fluid to be sterilized will dictate the polymer to be used. The listing below is intended to serve only as a guide for the selection of membrane filters for a particular application. The filter manufacturer should be consulted before making a final choice.

Fluid	Polymer
Aqueous	PVF, MCE
Oil	PVF, MCE
Organic solvents	PVF, PTFE
Aqueous, extreme pH	PVF
Gases	PVF, PTFE

Figure 40-20 is an example of a sterilizing filtration system commonly used in the pharmaceutical industry.

Positive pressure commonly is used in sterilizing filtrations. It has the following advantages over vacuum: it provides higher flow rates, integrity testing is easier, and it avoids a negative pressure on the downstream (sterile) side of the filtrate, thus precluding contamination. Membrane filters are sterilized readily by autoclaving, by *in-situ* steaming, or by using ethylene oxide.

In addition to their use in the pharmaceutical industry, membrane filters are used in many applications in the hospital pharmacy. The membrane filters commonly used in these applications are small disposable units. Examples of these are shown in Figures 40-21 and 40-22. Typical applications for membrane filters in hospital pharmacies include sterilization of intravenous (IV) admixtures and hyperalimentation solutions, sterilization of extemporaneously compounded preparations, sterility testing of admixtures, as well as in direct patient care (see Chapter 42).

Radiation Sterilization

The retail or hospital pharmacist probably has little opportunity to use radiation sterilization. However, they should be aware that many of the products sold in stores and used daily in hospitals are sterilized by this technology. Products such as contact lens solutions, bandages, baby bottle nipples, and teething rings (the kind containing water/gel) are a few exam-

Figure 40-20. An example of a process filtration system in a pharmaceutical plant (courtesy, Millipore).

ples of the everyday type of product encountered in a pharmacy. Several drugs, including some anticancer drugs, also are terminally sterilized using gamma radiation.

The hospital pharmacist is likely to encounter the use of gamma or X-ray treatment of blood to eliminate white blood cells in host-versus-graft reactions following transplant surgery. The serum used for tissue cultures is frequently ster-
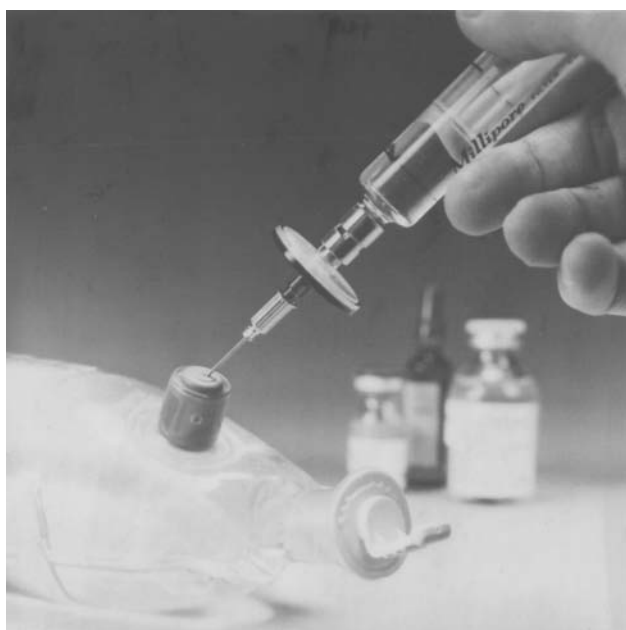
Figure 40-21. Intravenous additive filtration using a small disposable membrane filter (courtesy, Millipore).

Figure 40-22. Intravenous additive filtration and sterility testing. Both procedures employ membrane filtration (courtesy, Millipore).

ilized with gamma radiation to eliminate viruses, virus-like particles, and mycoplasms.

The pharmaceutical industry historically has relied on steam, dry heat, ethylene oxide gas, filtration, and chemical processes to meet sanitization or microbial load reduction requirements. Sterilization by radiation may employ either electromagnetic radiation or particle radiation.

Electromagnetic radiation, composed of photons of energy, includes ultraviolet, gamma, X-, and cosmic radiation. Gamma radiation, emitted from radioactive materials such as Cobalt-60 or Cesium-137, is the most frequently used source of electromagnetic radiation. Of these two, only Cobalt-60 is used in the large industrial irradiator (Fig 40-23). Cesium-137 is used in blood irradiators.

Particulate or corpuscular radiation includes a formidable list of particles. The only one that currently is being employed for sterilization is the electron. These electrons are machine generated using the technique illustrated in Figure 40-24; Figures 40-25 and 40-26 illustrate two methods of presenting products to a commercial electron-beam sterilizer.

Radiation-processing technology, and its application in the manufacture of pharmaceuticals, is being investigated more actively now than at any other time. This renewed interest is in part due to the development of aseptic and barrier technology, as well as an overall improvement in the environment in which pharmaceuticals are manufactured.

In the past the use of a radiation dose of 25 kGy was required to ensure that all viable microbes had been inactivated, and that a SAL of 10^{-6} was achieved. This level of radiation

proved detrimental to many pharmaceuticals. With the advent of clean rooms, and aseptic and barrier technologies, the microbial environment has improved dramatically. No longer are spores or even the number of organisms as daunting. It is more appropriate now to determine the resistance of the bioburden to radioaction and to tailor the minimal sterilization dose to meet the most resistant strain of the bioburden. In this way many more drugs and other products are capable of being sterilized terminally. This provides an SAL of 10^{-6} or greater, depending upon the microorganism.

The increased use of radiation processing to sterilize medical devices has led to the development of more efficient and economical irradiation equipment and processes. It also has generated new scientific data. The positive experience of the medical-device industry should be a *signpost* for the pharmaceuticals industry.

Several pharmaceutical raw materials and finished products are being sanitized/sterilized successfully with gamma radiation. Although it is possible to use electron beam radiation, we are presently unaware of any pharmaceuticals being treated using this technology. This should not preclude others from investigating its potential. The superior penetrating ability of gamma radiation provides the edge for this technology in this application.

HOW RADIATION KILLS MICROORGANISMS

The principles of sterilization by irradiation have been known since the early 1940s. Basically, charged particles or electromagnetic radiation interact with matter to cause both ionization and excitation. Ionization results in the formation of ion pairs, comprised of ejected orbital electrons (negatively charged) and their counterparts (positively charged). Charged particles such as electrons interact directly with matter causing ionization, whereas electromagnetic radiation causes ionization through various mechanisms that result in the ejection of an orbital electron with a specific amount of energy transferred from the incident gamma ray. These ejected electrons then behave similarly to machine-generated electrons in ionization reactions. Thus, both particle and electromagnetic radiation are considered as ionizing radiation and differ from ultraviolet radiation in this respect.

Ionizing radiation kills or inactivates microorganisms through the interaction of the ion pairs or excitations altering the molecular structure or spatial configuration of *biologically active* macromolecules. In particular, those involved in cell replication are most critical. It can do this in two ways. The first is to deposit energy directly in a bond of the macromolecule. This can cause a rearrangement of its structure, altering or destroying its normal function. The second is to generate free radicals, primarily from the water contained within the cytoplasm. The free radicals thus generated react with the macromolecules to subvert their normal function. In either case the result is the loss of reproductive capability of the microorganism.

The number of microorganisms inactivated by a given radiation dose is a statistical phenomenon. It depends upon the sensitivity of the biologically active macromolecule(s) to alteration (denaturation), the number of alterations elicited within the cell and the ability of the cell to repair these alterations. Different microorganisms have different capabilities to withstand or repair such alterations. This sensitivity is referred to as the D_{10} value. The size of the microorganism, its state of hydration, and the presence or absence of radical scavengers affect the outcome of exposure to ionizing radiation.

The ability of gamma radiation to inactivate microorganisms has been well documented. New documentation relating to viruses or new strains/reclassifications of microorganisms is being added continually. The major benefit of using radiation sterilization as the terminal step in the manufacturing process, as opposed to autoclaving or dry heat methods, is the minimal product degradation usually observed with this technology.
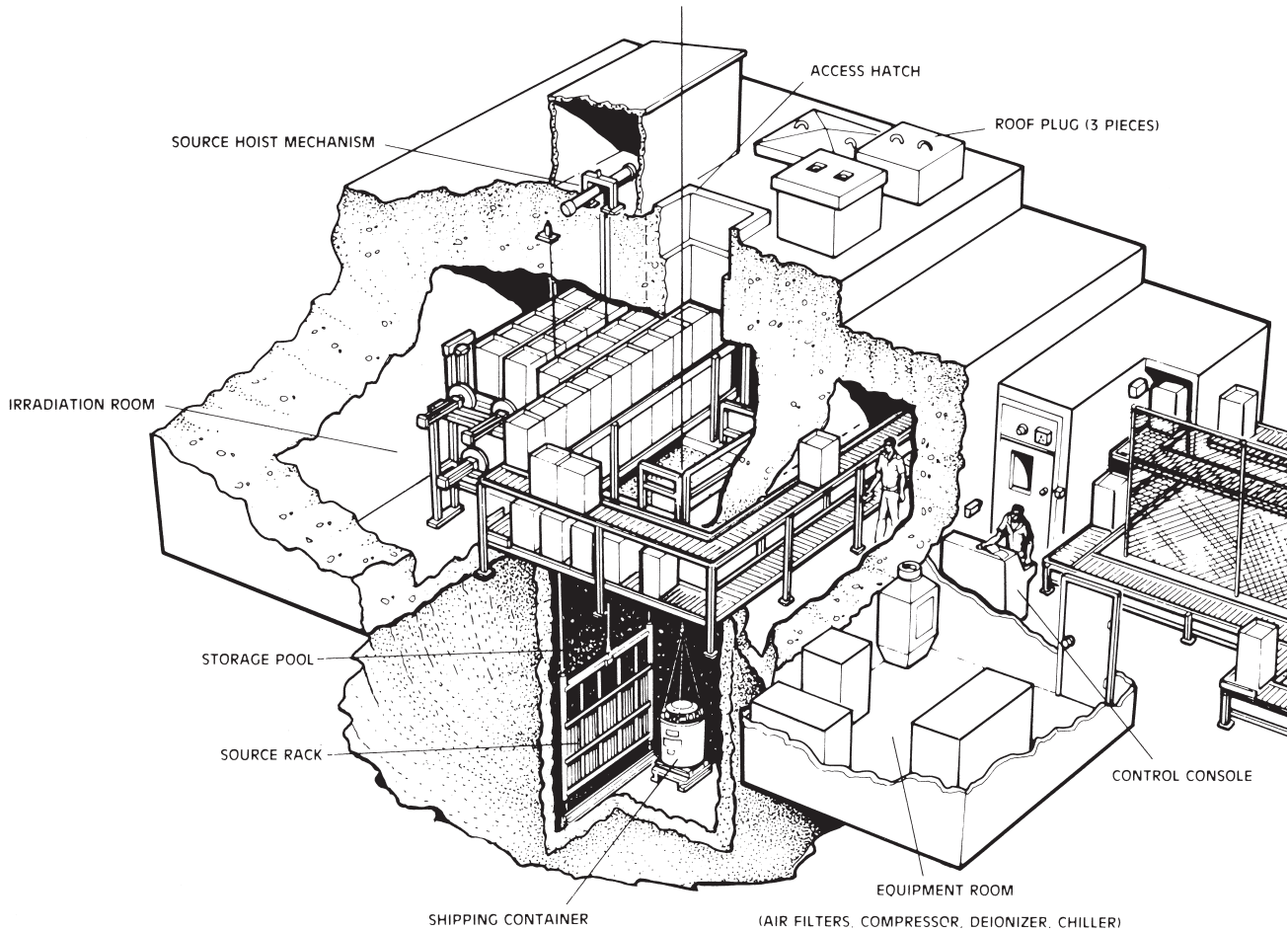
Figure 40-23. Tote box irradiator: automatic (courtesy, Nordion Intl).

The process has been in use in the medical device industry for over 25 years. Ample evidence as to its efficacy exists in scientific literature. Materials and processes have been developed to reduce the impact of radiation on the product. Some materials, such as Teflon and polypropylene, are severely degraded by radiation and must be avoided. It is the intent of this update to present some of the process developments that will facilitate the use of this technology for the terminal sterilization of pharmaceutical products. It also will assist those wishing to improve the microbial quality of raw materials entering the manufacturing process. Clean materials reduce the bioburden levels present in a clean room facility.

Sterilization by ionizing radiation requires consideration of the minimum and maximum doses (or the amount of radiation that is absorbed by the material), the energy level available (which along with the bulk density of the material will determine the depth of penetration), and the power output available (which determines the rate at which the dose can be applied).

The unit of absorbed dose is the Gray (Gy), where 1 Gy $= 1$ joule/kg, independent of the nature of the irradiated substance.
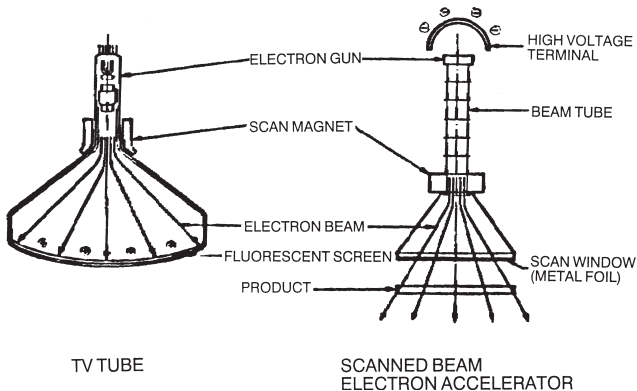
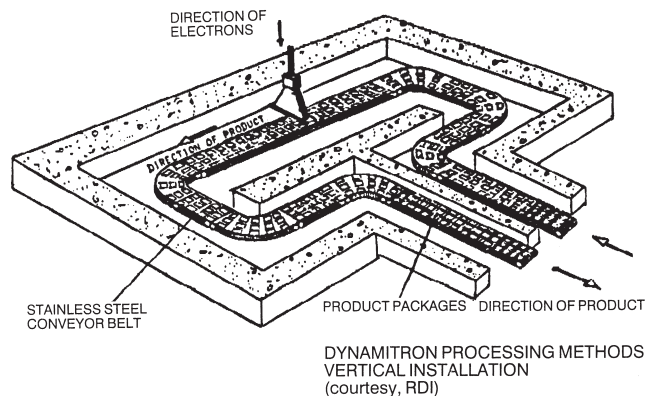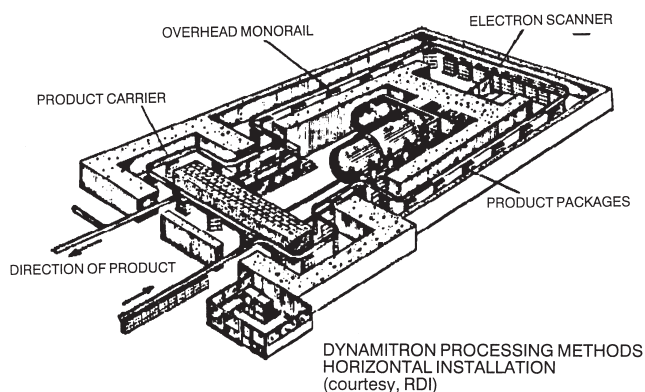**Figure 40-24.** To produce an electron (courtesy, RDI).

Figure 40-25. Dynamitron processing methods: vertical installation (courtesy, RDI).

Figure 40-26. Dynamitron processing methods: horizontal installation (courtesy, RDI).

Sterilization doses, for convenience, are predominantly expressed in kilogray (kGy).

Many investigators have studied the relative resistance of microorganisms to sterilization by radiation. The consensus is that vegetative forms are most sensitive, followed by molds, yeasts, viruses and spore-formers. While historical practice has been to use 15 to 25 kGy, today the minimal sterilization dose is more closely tailored to the resistance of the bioburden. It is not unusual to use doses as low as 2 to 8 kGy. The use of the AAMI/ISO or EN standards is highly recommended.

Modern gamma sterilization facilities used by pharmaceutical and medical device companies generally hold up to 4 MCi of Cobalt-60. The largest facility holds 12 MCi. Figure 40-23 shows a schematic of a modern Cobalt-60 radiationsterilization facility.

Two types of electron accelerators are used in sterilization: alternating-current machines with ranges up to 50 kW of power and 5 to 12 meV of energy, and direct-current machines with ranges of 30 to 200 kW and 0.5 to 5 meV. These machines generate electrons at high voltage, accelerate the electrons, and project them into the product to be sterilized. The greater the machine power (kW), the more electrons can be generated per unit time. The higher the energy (meV), the greater the penetration capability of the electrons into the material to be sterilized.

Ultraviolet (UV) Radiation

Artificially produced UV radiation in the region of 253.7 nm has been used as a germicide for many years. Although UV radiation often is used in the pharmaceutical industry for the maintenance of aseptic areas and rooms, it is of limited value as a sterilizing agent.

Inactivation of microorganisms by UV radiation is principally a function of the radiant energy dose, which varies widely for different microorganisms. The primary mechanism of microbial inactivation is the creation of the thymidine dimers in DNA, which prevents replication. Vegetative bacteria are most susceptible, while bacterial spores appear to be 3 to 10 times as resistant to inactivation and fungal spores may be 100 to 1000 times more resistant. Bacterial spores on stainless-steel surfaces require approximately 800 μW min/cm^2 for inactivation. By comparison, the black spores of *Aspergillus niger* require an exposure of over 5000 μW min/cm^2. Even with an adequate dose, however, the requirements for proper application of germicidal UV radiation in most pharmaceutical situations are such as to discourage its use for *sterilization* purposes. On the other hand, as an ancillary germicidal agent, UV radiation can be useful.

When using UV radiation, it is very important that lamps be cleaned periodically with alcohol and tested for output: also its use requires that personnel be properly protected; eye protection is particularly important.

The principal disadvantage to the use of germicidal UV radiation is its limited penetration—its 253.7 nm wavelength is screened out by most materials, allowing clumps of organisms, and those protected by dust or debris, to escape the lethal action. The use of UV radiation as a sterilizing agent is not recommended unless the material to be irradiated is very clean and free of crevices that can protect microorganisms. Many organisms are capable of repairing the UV-induced DNA damage using photoreactivation (light repair) and dark repair.

PULSED LIGHT

Recently, high-intensity visible light has been developed to a level that allows it to be used for certain sterilization applications. The advantages include extremely short exposure times (eg, 2 to 3 pulses of a few seconds) and relative ease in shielding the operations to provide operator safety. It can be used for surface sterilization and certain terminal sterilization applications. This is limited to packaging materials that are transparent to the wavelengths used. It is applicable for certain plastic materials, but not for Type I glass. This technique requires additional study, but has been shown to be effective against all organisms studied thus far.

Aseptic Processing

Although not actually a sterilization process, aseptic processing is a technique frequently used in the compounding of prescriptions or commercial products that will not withstand sterilization but in which all of the ingredients are sterile. In such cases, sterility must be maintained by using sterile materials and a controlled working environment. All containers and apparatus used should be sterilized by one of the previously mentioned processes and such work should be conducted only by an operator fully versed in the control of contamination. The use of laminar-airflow devices or barrier technology for aseptic processing is essential.

With the availability of sterile bulk drugs and sterilized syringe parts from manufacturers, the purchase of several pieces of equipment permits pharmacies to produce filled sterile unit-dose syringes with minimum effort. The equipment needs have been described in a paper by Patel and associates.[25] Figure 40-27[25] illustrates this system.

PACKAGING

Following exposure of a product to a well-controlled sterilization treatment, the packaging material of the product is expected to maintain sterility until the time of use. Packaging must be durable, provide for permanent-seal integrity, and have pore sizes small enough to prevent entry of contaminants. Obviously, the packaging must be compatible with the method of sterilization.

The package design is important if the contents are to be removed without recontamination. Tearing of plastics or paper can be tempered by coatings, and sealed containers should be tested carefully to ensure retention of sterility at the time of use.

If sterile material passes through many hands, it is important to provide a tamperproof closure to indicate if the container has been opened inadvertently. These four features—compatibility with sterilization, proven storage protection, ease of opening, tamper-proofing—are highly desirable characteristics of medical packaging.

For hospitals and pharmacies, there are a wide variety of woven reusable materials or nonwoven disposable materials that provide acceptable sterile barriers and are offered by major packaging suppliers. These suppliers normally conduct extensive programs to ensure the ability of the material to maintain sterility. Both hospitals and industry have guidelines and accepted practices for sterile-product packaging.[5]
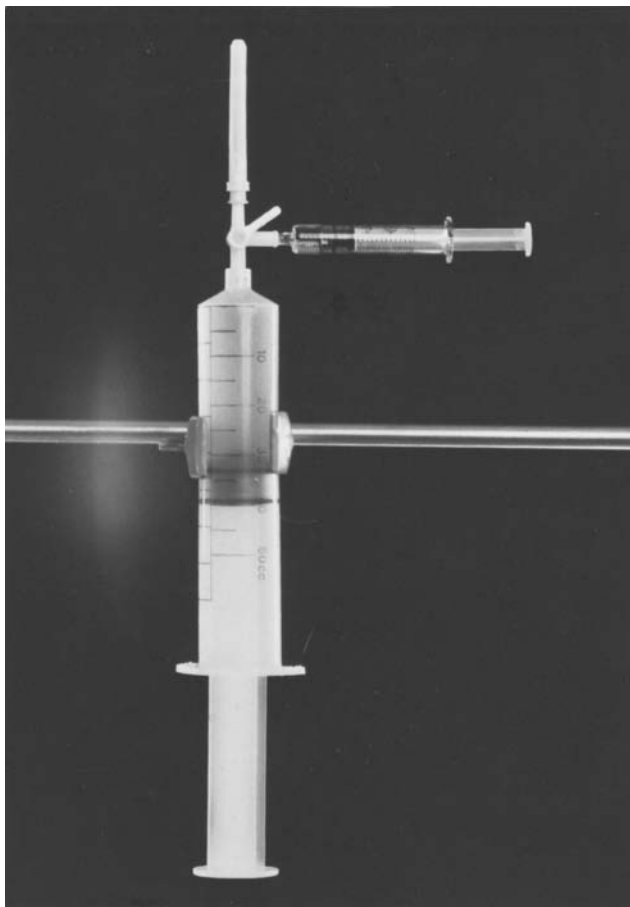
Figure 40-27. Unit-of-use system for sterile injectable medication.[19]

A review of the principles of sterile-material packaging by Powell[26] discusses the suitability of packaging materials for various sterilization methods, including resistance to bacteria, types of openings, strength of packaging, testing of packaging, and types of packaging. These topics also are discussed in Chapter 54.

UNIDIRECTIONAL AIRFLOW

Unidirectional airflow equipment is essential for proper performance of sterility tests and aseptic filling or assembling operations. These procedures require exact control over the working environment, but while many techniques and different types of equipment for performing these operations have been used over the years, unidirectional airflow devices are superior to all other environmental controls.

The unidirectional airflow procedure for producing very clean and dust-free areas was developed in 1961. In a unidirectional airflow device the entire body of air within a confined area moves with in one direction with uniform velocity along parallel flow lines. By employing prefilters and high-efficiency bacterial filters, the air delivered to the area essentially is sterile and sweeps all dust and airborne particles from the chamber through an open side. The velocity of the air used in such devices is generally 90 fpm ± 20%. Unidirectional airflow devices that deliver the clean air in a vertical, horizontal, or curvilinear fashion are available. The devices can be in the form of rooms, cabinets, or benches. For a comprehensive discussion of the biomedical application of unidirectional airflow the reader is referred to Runkle and Phillips.[27]

Each unidirectional airflow cabinet or bench should be located in a separate, small, clean room having a filtered air supply. The selection of the type of cabinet will depend on the oper-

Figure 40-28. Sterility testing of plastic disposable syringes in a horizontal laminar-airflow bench (courtesy, Becton Dickinson & Co).

ation itself. For most sterility-testing operations, horizontal unidirectional airflow units appear to be superior to vertical-flow hoods because the air movement is less likely to wash organisms from the operator's hands or equipment into the sterility test media. Figure 40-28 shows the sterility testing of syringes in a horizontal unidirectional airflow hood. Figure 40-29 shows the design of a typical horizontal, unidirectional airflow hood.

The major disadvantage of the horizontal unidirectional airflow units is that any airborne particulate matter generated in the units is blown directly into the room and against the working personnel. In situations where infectious material is involved, or where one must prevent contamination of the environment with a powder or drug, the use of specifically designed vertical, recirculating unidirectional airflow units is recommended. Units are available that do an excellent job of providing both product and personnel protection. Such a unit is shown in Figure 40-30.

To achieve maximum benefit from unidirectional airflow, it is important first to realize that the filtered airflow does not itself remove microbial contamination from the surface of objects. Thus, to avoid product or test contamination, it is necessary to reduce the microbial load on the outside of materials used in sterility testing. Unidirectional flow will do an excellent job of maintaining the sterility of an article bathed in the airflow; however, to be accurate, the sterility-testing, or product-assembly procedure must create the least possible turbulence within the unit. Moreover, an awareness of the turbulent air patterns created by the operation is necessary to avoid performing critical operations in turbulent zones. To illustrate how effectively airborne particles are washed from an environment by laminar airflow, Figure 40-31 shows the distance that particles of various sizes will travel horizontally before falling 5 ft in a cross-flow of air moving at 50 fpm.

Unidirectional airflow clean benches should supply Class 100 air as defined in Federal Standard 209B.[28] They should be certified to this standard when installed and then tested periodically. An air velocimeter should be used at regular intervals to check the airflow rates across the face of the filter. Smoke tests are use-
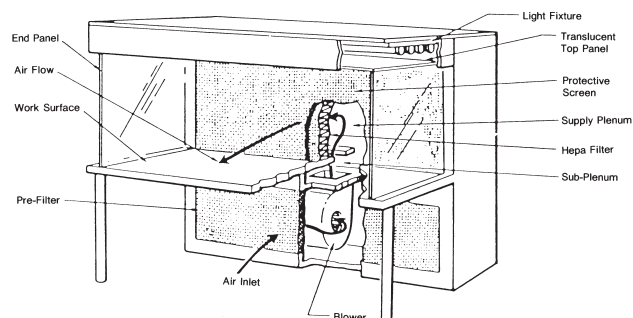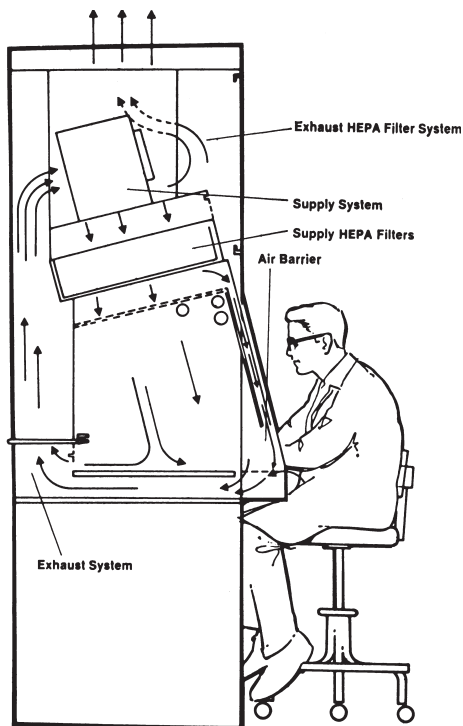

Figure 40-29. Horizontal laminar-airflow hood.

Figure 40-30. Sketch of a biological cabinet with vertical, recirculating laminar-airflow and HEPA-filtered exhaust. HEPA-filtered air is supplied to the work area at 90 fpm 20%. Airflow patterns in combination with a high-velocity curtain of air form a barrier at the front access opening that protects both the work and the worker from airborne contamination (courtesy, Bioquest).

ful in visualizing airflow patterns and a particle analyzer can be used to check the quality of the air. Filter efficiency testing determines the validity of the filter and its seal using a smoke (mean particulate diameter of 0.3 μm) and a light-scattering aerosol photometer. The smoke, at a concentration of 80 to 100 mg/L, is introduced to the plenum of the unit and the entire perimeter of the filter face is scanned with the photometer probe at a sampling rate of 1 ft³/min. A reading of 0.01% of the upstream smoke concentration is considered a leak.

In addition to the routine airflow measurements and filter-efficiency testing, biological testing should be done to monitor the effectiveness of laminar-airflow systems. Microbial air sampling and agar-settling plates are useful in monitoring these environments. Phillips evaluated horizontal laminar-flow hoods by tabulating the number of *false positives* appearing in sterility-test media over a period of time. These results (Table 40-3) showed very low numbers of *false positives*.
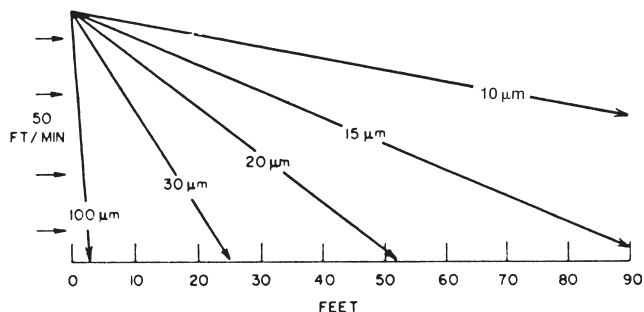
Figure 40-31. Distance traveled by particles settling from a height of 5 ft.

Table 40-3. False Positives Occuring in a Laminiar-Flow Hood[26]

PRODUCT	NO. OF UNITS STERILITY TESTED	NO. OF FALSE POSITIVES	% FALSE POSITIVES
Syringes	9793	2	0.02
Needles	4676	2	0.04
Misc	306	0	0

See Figure 40-29 for laminar-flow hood.

TESTING

After sterilization, there are several techniques for determining whether the particular lot of material is sterile. The only method for determining sterility with 100% assurance would be to run a total sterility test, that is, to test every item in the lot.

Representative probabilities are shown in Tables 40-4 and 40-5 to illustrate more specifically how low levels of contamination in treated lots of medical articles may escape detection by the usual sterility-test procedures. The data are calculated by binomial expansion, employing certain assumed values of percent contamination with large lot sizes (greater than 5000) and including standard assumptions with regard to the efficiency of recovery media and so on.

In Table 40-4 the probability data are calculated for lots with various degrees of assumed contamination when 10 random samples per lot are tested. For example, a lot that has one in each 1000 items contaminated (0.1% contamination) could be passed as satisfactory (by showing no positive samples from 10 tested) in 99 tests out of 100. Even at the 10% contamination level, contamination would be detected only two out of three times.

Table 40-5 shows the difficulty in attempting to improve the reliability of sterility tests by increasing sample size. For contamination levels as low as 0.1%, increasing the sample size from 10 to 100 has a relatively small effect in improving the probability of accepting lots. Even a sample size of 500 would result in erroneously accepting a lot 6 times out of 10. On the other hand, with a lot contaminated to the extent of 10%, by testing 100 samples the probability of acceptance of the lot would be reduced to a theoretical zero.

The information in Table 40-5 may be viewed in another way. If, for the probability values shown for each different sample size, the value that approximates the 95% confidence level ($P = 0.05$) is selected, it is clear that using 20 samples only will discriminate contamination levels of 15% or more. If the 20 tubes show no growth, the lot could, of course, be sterile but there would be no way of knowing this from the test. From such a test it could be stated only that it is unlikely that the lot would be contaminated at a level higher than 15%. It is clear from these data that product sterility testing is a poor method of validating sterilization procedures.

The USP provides two basic methods for sterility testing. One involves the direct introduction of product test samples into culture media; the second involves filtering test samples through membrane filters, washing the filters with fluids to remove inhibitory properties, and transferring the membrane

Table 40-4. Probabilities for Sterility Testing of Articles With Assumed Levels of Contamination

"TRUE" % CONTAMINATION	PROBABILITY OF DESIGNATED POSITIVES OUT OF 10 SAMPLES TESTED			
	0	2	5	10
0.1	0.990	(Total = 0.010)		
1.0	0.904	0.091		
5.0	0.599	0.315		
10.0	0.349	0.387	0.001	
30.0	0.028	0.121	0.103	
50.0	0.001	0.010	0.246	0.001

Table 40-5. Relationship of Probabilities of Acceptance of Lots of Varying Assumed Degrees of Contamination to Sample Size

NUMBER OF SAMPLES TESTED (*n*)	PROBABILITY OF NO POSITIVE GROWTH "TRUE" % CONTAMINATION OF LOT					
	0.1	1	5	10	15	20
10	0.99	0.91	0.60	0.35	0.20	0.11
20	0.98	0.82	0.36	0.12	0.04	0.01
50	0.95	0.61	0.08	0.007		
100	0.91	0.37	0.01	0.00		
300	0.74	0.05				
500	0.61	0.01				

aseptically to appropriate culture media. Test samples may be sterilized devices that simply are immersed aseptically into the appropriate culture-broth washings of the sterile object with sterile diluent, or dilutions of sterile materials. The USP recommends three aqueous diluting fluids for sterility tests while the Antibiotic Regulations list four; all are nontoxic to microorganisms. In the case of petrolatum-based drugs, a nonaqueous diluting fluid is required.

Many studies have been conducted to find the minimum number of culture media that will provide the greatest sensitivity in detecting contamination. Internationally recognized experts and bodies now recommend the use of two culture media: Soybean-Casein Digest Medium, incubated at 20° to 25°C, and Fluid Thioglycollate Medium, incubated at 30° to 35°C. The time of incubation specified usually is 7 days for the membrane filtration method and 7 to 14 days for the direct-inoculation method, depending on the method of sterilization. The requirements are described in detail in the USP.

The preferred method of verifying sterility is not by testing sterilized materials but by the use of biological indicators. This is not possible, however, when products are sterilized by filtration and filled aseptically into their final containers, as is the case with such important drugs as antibiotics, insulin, or hormones. The indicators generally are highly resistant bacterial spores present in greater numbers than the normal contamination of the product and with equal or greater resistance than normal microbial flora in the products being sterilized. Various properties of commercially available bacterial spores have been recommended for specific methods of sterilization based on unique resistance characteristics.

Commonly accepted species of bacteria used for biological indicators are shown in Table 40-6. Other species can be employed, probably without serious impact on the validity of sterility interpretation, so long as the prime requirements of greater numbers and higher resistance, compared to material contamination characteristics, are maintained.

Included with the materials being sterilized, biological indicators are imbedded on either paper or plastic strips or are inoculated directly onto the material being sterilized. Obviously, the indicator has greater validity in verifying sterility if it is located within product spaces that are the most difficult to sterilize. For example, in the case of a syringe, the location of a paper strip or inoculation of spores between the ribs of the plunger stopper is recommended.

The use of isolators (barrier technology) for processing materials is discussed in the section on advanced aseptic process-

Table 40-6. Species of Bacteria Used as Biological Indicators

METHOD OF STERILIZATION	BACTERIAL SPECIES
Moist heat	*B stearothermophilus*
Dry heat	*B subtilis*
Ethylene oxide	*B stearothermophilus*
Radiation	*B pumilus, B stearothermophilus, B subtilis*

Figure 40-32. Stainless steel half-suit isolator (courtesy, Laminar Flow).

ing. The first widespread use of these modern *glove boxes* in the pharmaceutical industry was in sterility testing. As filling speeds became faster, batches became larger. This, coupled with more expensive drug substances, created the need to avoid false-positive sterility tests. Even with laminar-flow hoods becoming widely used, the large number of manipulations carried out by people, created a significant chance for contamination through the testing procedure.

Government standards for SAL basically eliminated the possibility to repeat sterility tests. This means that batches that fail for any reason cannot be released. They are only useful for investigation of potential contamination hazards. Industry in turn needed more assurance that the product was indeed not sterile and the test was valid. This led to the development of more sophisticated isolation units. Figure 40-32 shows a stainless-steel half-suit system that is typical of those used in sterility testing. The units can be *docked* to a sterilizer, which eliminates the possibility of contamination during transfer of materials to the test area. The units can be sterilized using vaporized hydrogen peroxide. The exterior of any test materials required to be transferred into the units also can be sterilized in this manner. Validation of these steps allows one to virtually eliminate false-positive test results. Most manufacturers have adopted this technique and have agreed to a policy of essentially no sterility retests. Only where obvious system breakdowns can be shown to have led to failures will a retest be considered.

ADVANCED ASEPTIC PROCESSING

Isolator technology also is being used with increasing frequency in the processing of sterile products and associated packaging materials. This is driven by the same need to minimize human intervention and thus increase dramatically the assurance of sterility (SAL). The minimization of people was expanded throughout the 1990s with the advent of more widespread use of form, fill, and seal technology. This involved the on-line molding, filling, and sealing of plastic bottles containing sterile products. The technology was housed in rigid walled areas and product was supplied through filters and sterilized in place, at the last possible area before filling.

While form, fill, and seal is a technology on its own, it did lead to the recognition that by updating significantly the older concepts of *glove boxes,* one could dramatically affect the sterility assurance of an aseptic process. People contribute the

largest percentage of the contamination risk. By minimizing their interaction, the probability of nonsterility is greatly reduced.

Glove boxes were not designed to support modern (and especially more automated) operations. This may explain why they did not become popular as aseptic processing units. Usage began to expand only when the need for increased assurance drove designers to develop ergonomically designed isolators.

More recently, the manufacturers of high-speed processing equipment have begun to redesign their machinery in line with the principles of isolator units. Because the mechanics of these machines have been proved to be very reliable and require very little human intervention, the timing seems to be correct for such modifications. Isolator units are relatively inexpensive also. They have allowed for aseptic processing without the construction of large processing areas, sterile suites or gowning areas. The development of relatively safe sterilization methods, such as vaporized hydrogen peroxide (with or without atmospheric steam) also has allowed the technology to become more viable for widespread use. Prior to this, the use of toxic (and sometimes corrosive) materials was required. This limited their use to more sophisticated operators, who were able to afford the resources required to build and maintain facilities for their use.

TRAINING

It is desirable that personnel involved with sterilization or aseptic processing be instructed in the basic behavior of microorganisms. This would include the differentiation of vegetative, spore-forming, and slow-growing life forms such as molds and yeasts. This would allow those being trained in the operations to understand the reasons for many of the restrictions necessary to carry out these processes. It is imperative that each person involved in these operations be instructed in two main areas.

Safety is the first and foremost area of concentration for a training program. Each of the pieces of equipment and processes described above have unique hazards associated with them. The operators must be made to understand the dangers of steam under pressure and exposure to gaseous sterilants prior to their neutralization.

The equipment design and installation should undergo safety reviews prior to its general operation. This review for potential hazards must be done by highly trained individuals and should include computer control and piping systems. It is important that the equipment fail (should a failure occur) in a manner that is safe to the operators. Valves should fail in a way to vent pressure to some safe area and/or gases to a relatively safe, unoccupied site.

The second major area of training involves gowning for entry into the sterile areas and subsequent performance of aseptic operations. Personnel must be instructed in proper gowning techniques so that they do not contaminate the exterior of garments and gloves during the process. Gowning areas should be supplied with full-length mirrors so that personnel can verify that all areas of their body have been covered fully and properly prior to entering a sterile work area. Recent trends indicate that gowning training should be followed by personnel monitoring with contact plates containing growth media. This allows one to verify the effectiveness of the training and, should growth occur, one can use this growth as a training tool to emphasize the importance of careful attention to detail during the gowning process. Because these plates require incubation, one does not allow operators to enter the sterile area until the results of these tests have been collected and reviewed with the candidate.

Continuing with the above approach, those performing aseptic operations require additional training and subsequent verification. This principle of competency-based training (ie, verifying the capabilities of those being trained) is necessary to ensure that the operators have developed the skills to carry out these vital operations while minimizing the risk of contamination. Again, it allows for constructive feedback to those who have not yet become fully accomplished in the techniques. It is prudent to reinforce these skills periodically through refresher sessions, and reverification of the skills. It has become standard practice to do unannounced spot-checks of the gloves and gowns of aseptic operators. This practice helps to maintain a level of vigilance, with regard to proper gowning and operating technique.

REFERENCES

1. *Medical Device Sterilization Monographs* (Rep Nos 78-4.13 and 78-4.11). Washington, DC: Health Industry Manufacturers Association, 1978.
2. Block SS, ed. *Disinfection, Sterilization and Preservation,* 3rd ed. Philadelphia: Lea & Febiger, 1983.
3. *Steam Sterilization and Sterility Assurance, Good Hospital Practice* (AAMI Recommended Practice, ST.1-1980). Arlington VA: Assoc Adv Med Instrum, 1980.
4. *Guideline for Industrial Ethylene Oxide Sterilization of Medical Devices* (AAMI Recommended Practice, OPEO-87). Arlington, VA: Assoc Adv Med Instrum, 1987.
5. *In-Hospital Sterility Assurance—Current Perspectives, Aseptic Barrier Evaluation, Sterilizer Processing, Issues in Infection Control and Sterility Assurance* (AAMI Technol Assess Rep No 4-82). Arlington, VA: Assoc Adv Med Instrum, 1982.
6. *Hospital Steam Sterilizers* (Am Natl Std, ANSI/AAMI ST8-1982). Arlington, VA: Assoc Adv Med Instrum, 1983.
7. *Process Control Guidelines for Gamma Radiation Sterilization of Medical Devices* (AAMI Recommended Practice, RS-3/84). Arlington, VA: Assoc Adv Med Instrum, 1984.
8. *Performance Evaluation of Ethylene Oxide Sterilizers—Ethylene Oxide Test Packs, Good Hospital Practice* (AAMI Recommended Practice, EOTP-2.85). Arlington, VA: Assoc Adv Med Instrum, 1985.
9. *Biological Indicators for Saturated Steam Sterilization Processes in Health Care Facilities* (Am Natl Std, ANSI/AAMI ST 19-1985). Arlington, VA: Assoc Adv Med Instrum, 1986.
10. *Good Hospital Practice: Steam Sterilization Using the Unwrapped Method (Flash Sterilization)* (AAMI Recommended Practice, SSUM-9/85). Arlington, VA: Assoc Adv Med Instrum, 1986.
11. Johnson J. *Vaporized Hydrogen Peroxide Sterilization of Freeze Dryers.* ISPE Ann Meeting, Panama City, FL, 1993.
12. Lysfjord JP, et al. *The Potential For Use of Steam at Atmospheric Pressure to Decontaminate or Sterilize Parenteral Filling Lines Incorporating Barrier Isolation Technology.* Spring Mtg of the PDA, Philadelphia, 10 Mar 1993.
13. Edwards LM. *Pharm Eng* 1993; 13(2):50.
14. Rosenblatt, et al. *Use of Chlorine Dioxide Gas as a Chemosterilizing Agent,* US Pat 4,504,422 (Scopas Technol Corp), 1985.
15. Knapp JE, Rosenblatt DH, Rosenblatt AA. *Med Dev Diag Ind* 1986; 8:48.
16. Kowalski JB, Hollis RA, Roman CA. In: Pierce G, ed. *Developments in Industrial Microbiology,* vol 29. Amsterdam: Elsevier, 1988, p 239.
17. Jeng DK, Woodworth AG. *Artif Organs* 1990; 14:361.
18. National Archives. *Federal Register* 40: Mar 14, 1975, p 11865.
19. 21 CFR 211.72.
20. *Microbiological Evaluation of Filters for Sterilizing Liquids,* vol 4, no 3. Washington, DC: Health Ind Manuf Assoc, 1981.
21. Leahy TJ, et al. *Pharm Technol* 1978; 2:65.
22. Dawson FW, et al. *Nordiska Foreningen for Renlighelsteknik och Rena Rum,* Goteborg, Sweden, 1981, p 5.
23. *Test for Determination of Characteristics of Membrane Filters for Use in Aerospace Liquids (Proposed Tentative Test Method).* Philadelphia: ASTM, June 1965.
24. Reti AR, et al. *Bull Parenteral Drug Assoc* 1977; 31:187.
25. Patel JA, Curtis EG, Phillips GL. *Am J Hosp Pharm* 1972; 29:947.
26. Powell DB. In: Phillips GB, Miller WS, eds. *Industrial Sterilization.* Durham, NC: Duke University Press, 1973, p 79.
27. Runkle RS, Phillips GB, eds. *Microbial Contamination Control Facilities.* New York: Van Nostrand-Reinhold, 1969.
28. *Clean Room and Work Station Requirements: Controlled Environment* (Fed Std No 209B). Washington, DC: USGPO, 24 Apr 1973.