PART FOUR

DOSAGE FORM DESIGN AND MANUFACTURE

This page intentionally left blank

21
Solutions

Michael Billany

CHAPTER CONTENTS

INTRODUCTION

An understanding of the properties of solutions, the factors that affect solubility and the process of dissolution is essential because of the importance of solutions in so many areas of pharmaceutical formulation. These basic topics are covered in Chapters 2 and 3 and it is recommended that they be read in conjunction with this chapter for a full understanding of principles involved in the formulation of solutions.

Ways of expressing solubility and definitions of terms are also explained in these earlier chapters, but it is worth reiterating that a solution is a homogenous one-phase system consisting of two or more components. The solvent, or mixture of solvents, is the phase in which the dispersion occurs, and the solute is the component which is dispersed as molecules or ions in the solvent.

In general the solvent is present in the greater amount, but there are several exceptions. For example, Syrup BP contains 66.7% w/w of sucrose as the solute in 33.3% w/w of water as the solvent.

For most pharmaceutical solutions the solvent system is likely to be liquid, and the solute will be either a liquid or a solid. Solid dispersions, in which both solute and solvent are solids, are used for the improvement of the bioavailability of poorly soluble drugs. The solute is present as a molecular dispersion and will therefore exhibit a fast rate of dissolution, owing to its very high specific surface area. In addition, the particles are already in a deaggregated and wetted state, and so there is little air adsorbed on the particle surfaces to inhibit dissolution. There may even be a slight increase in actual solubility.

Solutions of gases in liquids are characteristic of aerosols, in which the propellant gas is dispersed or dissolved in the solvent under pressure. On actuation of the valve mechanism the propellant ejects the product from the container. The propellant immediately evaporates, leaving the active agent in the form of tiny droplets or particles or within a foam structure.

ADVANTAGES AND DISADVANTAGES OF SOLUTIONS AS AN ORAL DOSAGE FORM

Although tablets and capsules are more widely used than liquid preparations for oral administration, the latter possess several advantages:

- Liquids are easier to swallow than solids and are therefore particularly acceptable for paediatric and geriatric use.
- A drug must usually be in solution before it can be absorbed. If it is administered in the form of a solution, the drug is immediately available for absorption. Therefore, the therapeutic response is faster than if using a solid dosage form, which must first disintegrate in order to allow the drug to dissolve in the gastrointestinal fluid before absorption can begin. Even if the drug should precipitate from solution in the acid conditions of the stomach, it will be in a sufficiently wetted and finely divided state to allow rapid absorption to occur.
- A solution is a homogenous system and therefore the drug will be uniformly distributed throughout the preparation. In suspension or emulsion formulations uneven dosage can occur as a result of phase separation on storage.
- Some drugs, including aspirin and potassium chloride, can irritate and damage the gastric mucosa, particularly if localized in one area, as often occurs after the ingestion of a solid dosage form. Irritation is reduced by the administration of a solution of a drug because of the immediate dilution by the gastric contents.

There are, however, several problems associated with the manufacture, transport, stability and administration of solutions:

- Liquids are bulky and therefore inconvenient to transport and store. If the container should break the whole of the product is immediately and irretrievably lost.
- The stability of ingredients in aqueous solution is often poorer than if formulated as a tablet or capsule, particularly if they are susceptible to hydrolysis. The shelf-life of a liquid dosage form is often much shorter than that of the corresponding solid preparation. Not only is the stability of the drug important, but also that of other excipients, such as surfactants, preservatives, flavours and colours. The chemical stability of some ingredients can, however, be improved by the use of a mixed solvent system. The inclusion of a surfactant above its critical micelle concentration can also improve chemical stability because the hydrolytic degradation of a material may be reduced by its solubilization within the micelles.
- Solutions often provide suitable media for the growth of microorganisms and may therefore require the incorporation of a preservative.

- Most liquid preparations are designed so that the normal dosage of the drug is present in 5 mL, or a multiple of 5 mL, of product. Accurate dosage depends on the ability of the patient to use a 5 mL spoon or a volumetric dropper.
- The taste of a drug, which is usually unpleasant, is always more pronounced when in solution than in a solid form. Solutions can, however, easily be sweetened and flavoured to make them more palatable.

CHOICE OF SOLVENT

Aqueous solutions

Water is the solvent most widely used as a vehicle for pharmaceutical products, because of its physiological compatibility and lack of toxicity. It possesses a high dielectric constant, which is essential for ensuring the dissolution of a wide range of ionizable materials. In some cases this property may be an advantage, but the lack of selectivity can be responsible for aqueous solutions containing unwanted substances such as inorganic salts and organic impurities. This is one reason why water is rarely used for the extraction of active constituents from vegetable sources.

Types of pharmaceutical water

For many preparations potable water can be used. This is water freshly drawn from the mains system and which is suitable for drinking. If this type is unobtainable, then a suitable, though more expensive, alternative is pharmacopoeial Purified Water BP, which has been freshly boiled and cooled immediately before use to destroy any vegetative microorganisms that might be present. Purified Water must, however, be used on all occasions where the presence of salts – often dissolved in potable water – is undesirable. Purified Water is normally prepared by the distillation or deionization of potable water, or by the process of reverse osmosis.

Water for Injections must be used for the formulation of parenteral solutions and is obtained by sterilizing pyrogen-free distilled water immediately after its collection. For the formulation of aqueous solutions of drugs, such as phenobarbitone sodium or aminophylline, which are sensitive to the presence of carbon dioxide, Water for Injections free from carbon dioxide must be used. Similarly, drugs which are liable to oxidation, such as apomorphine and ergotamine maleate, require Water for Injections BP free from dissolved air to be used.

These are both obtained from apyrogenic distilled water in the same way as before, but are then boiled for at least 10 minutes, cooled, sealed in their containers while excluding air, and then sterilized by autoclaving. For further details on parenteral solutions see Chapter 35.

Approaches to the improvement of aqueous solubility

Although water is very widely used for inclusion in pharmaceutical preparations, it may not be possible to ensure complete solution of all ingredients at all normal storage temperatures. Strongly ionized materials are likely to be freely soluble in water over a wide pH range. Similarly, weak acids and bases should be adequately soluble at favourable pHs. Even if in solution, it is still important to ensure that the concentration of any material is not close to its limit of solubility, as precipitation may occur if the product is cooled or if any evaporation of the vehicle should occur. For unionized drugs or for weak electrolytes at a pH that is unfavourable for extensive ionization, one or more of the following methods should be used to improve aqueous solubility.

Cosolvency The solubility of a weak electrolyte or non-polar compound in water can often be improved by altering the polarity of the solvent. This can be achieved by the addition of another solvent that is both miscible with water and in which the compound is also soluble. Vehicles used in combination to increase the solubility of a drug are called cosolvents, and often the solubility in this mixed system is greater than can be predicted from the material's solubility in each individual solvent.

Because it has been shown that the solubility of a given drug is maximal at a particular dielectric constant of any solvent system, it is possible to eliminate those solvent blends possessing other dielectric constants. In some cases, however, it has been shown that the chemical nature of the solvent system used may be of greater importance. A more detailed explanation of the effects of the molecular structure of the solute and the physicochemical properties of the solvent can be found in Chapter 2.

The choice of suitable cosolvents is somewhat limited for pharmaceutical use because of possible toxicity and irritancy, particularly if required for oral or parenteral use. Ideally, suitable blends should possess values of dielectric constant between 25 and 80. The most widely used system that will cover this range is a water/ethanol blend. Other suitable

solvents for use with water include sorbitol, glycerol, propylene glycol and syrup. For example, a blend of propylene glycol and water is used to improve the solubility of co-trimoxazole, and paracetamol is formulated as an elixir by the use of alcohol, propylene glycol and syrup. For external application to the scalp, betamethasone valerate is available dissolved in a water/isopropyl alcohol mixture.

For further details covering the suitability of different cosolvents see under their individual headings in this chapter.

pH control A large number of drugs are either weak acids or weak bases, and therefore their solubilities in water can be influenced by the pH of the system. A quantitative application of the Henderson–Hasselbalch equation will enable the solubility of such a drug in water at a given pH to be determined, provided its pK_a and the solubility of its unionized species are known (see Chapter 3). The solubility of a weak base can be increased by lowering the pH of its solution, whereas the solubility of a weak acid is improved by an increase in pH. Some compounds will accept or donate more than one hydrogen ion per molecule, and will therefore possess more than one pK_a value and so will exhibit a more complex solubility profile.

In controlling the solubility of a drug in this way, it must be ensured that the chosen pH does not conflict with other product requirements. For example, the chemical stability of a drug may also depend on pH, and in many cases the pH of optimum solubility does not coincide with the pH of optimum stability. This may also be true for other ingredients, especially colours, preservatives and flavours.

The pH of solutions for parenteral and ophthalmic use, for application to mucous membranes or for use on abraded skin must also be controlled, as extremes can cause pain, irritation and even tissue damage. This is particularly true for subcutaneous, intramuscular and intraspinal injections because the solutions will not be rapidly diluted after administration.

In some instances the bioavailability of drugs may be influenced by the pH of their solutions (Chapter 17), and changes in pH can also affect a preservative's activity by altering the extent to which it is ionized. Often a compromise must be reached during formulation to ensure that the stability and solubility of all ingredients, physiological compatibility and bioavailability are all adequate for the product's intended purpose.

The values of molar solubilities and dissociation constants of drugs that are reported in the literature, or determined during preformulation studies, are usually for the drug alone in distilled water. These values may differ in the final formulation owing to the presence of other ingredients. For example, the inclusion of cosolvents such as alcohol or propylene glycol with water will lower the dielectric constant of the vehicle and therefore increase the solubility of the unionized form of the drug. This lowering of the polarity of the solvent system will also reduce the degree of dissociation of the drug and increase its pK_a. As this effect will increase the concentration of the unionized (less soluble) species, an increase in the pH of the system may be necessary in order to maintain solubility.

It must be appreciated that maximum solubility may best be achieved by a judicious balance between pH control and concentration of cosolvent, and can be determined, as before, by the Henderson–Hasselbalch equation, substituting the new values both for pK_a and for the molar solubilities of the unionized species.

Suitable buffer systems for the control of pH are discussed later in this chapter, but care must be taken because the solubilities of sparingly soluble electrolytes can be decreased still further by the addition of a soluble electrolyte, should they contain a common ion. The opposite can be true if they do not possess common ions.

As solutions of non-electrolytes are not significantly affected by pH, other methods of improving their solubilities must be found.

Solubilization The solubility of a drug that is normally insoluble or poorly soluble in water can often be improved by the addition of a surface-active agent. These molecules form different types of micelles, ranging from simple spherical structures to more complex liposomes and liquid crystals. Details of their formation can be found in Chapter 6. This phenomenon of micellar solubilization has been widely used for the formulation of solutions of poorly soluble drugs. In aqueous systems, non-polar molecules will dissolve in the interior of the micelle, which consists of the lipophilic hydrocarbon moiety.

The amount of surfactant to be used for this purpose must be carefully controlled. A large excess is undesirable because of cost, possible toxicity and its effect on product aeration during manufacture. Excessive amounts may also reduce the bioavailability of a drug if it is strongly adsorbed within the micelle. An insufficient amount of surfactant, however, may not solubilize all the drug, or may lead to precipitation either on storage or on dilution of the product.

Reference to Fig. 6.15 will show that hydrophilic surfactants possessing HLB values above 15 will be particularly valuable as solubilizing agents.

The surfactant chosen must be non-toxic and non-irritant, bearing in mind its intended route of administration. It must also be miscible with the solvent system, compatible with the other ingredients, free from disagreeable odour and taste and be non-volatile.

Examples include the solubilization of fat-soluble vitamins such as phytomenadione using polysorbates. This enables their inclusion with water-soluble vitamins in the same aqueous-based formulation. For parenteral administration of these vitamins, a mixture of glycocholic acid and lecithin provides a mixed micelle system.

The solubility of amiodarone hydrochloride can similarly be improved, although this drug can exhibit autosolubilization at high concentrations.

The solubilization of iodine to produce iodophores is achieved by the use of macrogol ethers. These products exhibit several advantages over simple iodine solutions, including an improved chemical stability, reduced loss of active agent due to sublimation, less corrosion of surgical instruments and, in some cases, enhanced activity.

Polyoxyethylated castor oil is used as a solubilizing agent for a number of intravenous injections, including the immunosuppressant cyclosporin as an intravenous infusion. Care must be taken with this surfactant, as anaphylactic reactions are known to occur when it is injected.

Other drugs that have been solubilized include antibiotics such as griseofulvin, which has been formulated with cetomacrogol. Poloxamers (polyoxyethylene/polyoxypropylene copolymers), some of which are also suitable for parenteral administration, are used to maintain the clarity of solutions for oral use (Garcia Sagredo et al 1994). Lanolin derivatives have also been used for the solubilization of volatile and essential oils.

The solubility of phenolic compounds such as cresol and chloroxylenol, which are normally soluble in water up to 2% and 0.03% respectively, can be improved by solubilization with soaps. Lysol contains 50% cresol in an aqueous system by the use of the potassium soaps of oleic, linoleic and linolenic acids.

It may also be possible to combine the beneficial effects of solubilization and cosolvency in one formulation. A 5% chloroxylenol solution can be formulated by the inclusion of potassium ricinoleate. This soap is formed in situ by the reaction between potassium hydroxide and castor oil. Ethanol and terpineol are included as cosolvents.

To ensure that the optimum concentration of surfactant is chosen, a known weight or volume is added to each of a series of vials containing the solvent. Ensuring adequate temperature control, varying amounts of solubilizate (the material to be solubilized) are added to each vial in ascending order of concentration. The maximum concentration of drug that will form a clear solution with a given concentration of surfactant can be determined visually or by optical density measurement (Fig. 21.1(a)), and is known as the maximum additive concentration (MAC). This method can be repeated for different amounts of surfactant to enable a graph to be constructed of MAC against surfactant concentration, from which the optimum amount of solubilizing agent can be chosen for any required amount of drug (Fig. 21.1(b)).

Alternatively, a ternary-phase diagram can be constructed (Fig. 21.2) that will present a more comprehensive picture of the effects of solubilizate, surfactant and solvent concentrations on the physical characteristics of the system.

The three axes form the three sides of an equilateral triangle, each axis representing 0–100% of one of the components. Point A thus represents a formulation consisting of 50% solubilizate, 20% surfactant and 30% water. By plotting at each point a number

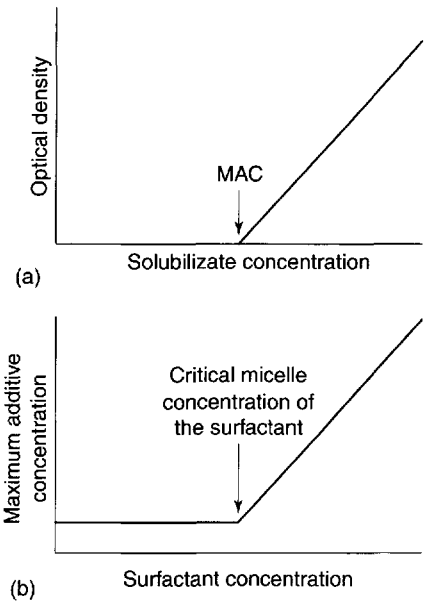


Fig. 21.1 (a) Graph of optical density of a solubilizate/surfactant/solvent system against solubilizate concentration showing the maximum additive concentration (MAC). (b) Determination of the MAC for a range of concentrations of a given surfactant will provide the data for this graph, enabling the optimum concentration of the surfactant to be chosen for the solubilization of a given concentration of active agent.

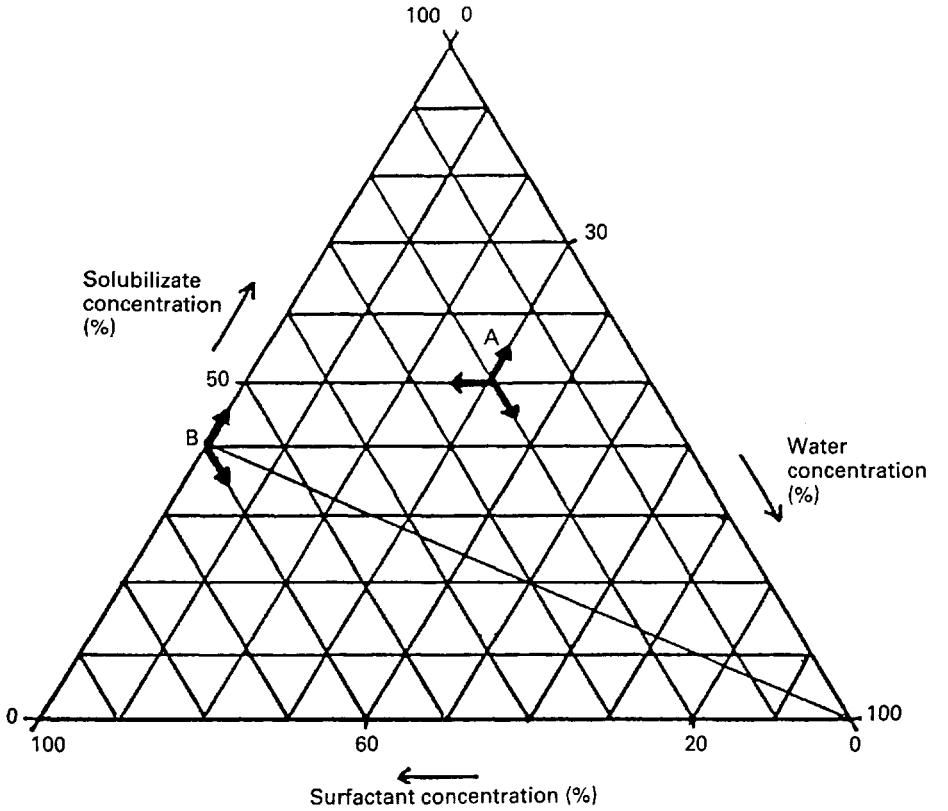


Fig. 21.2 Construction of a ternary-phase diagram.

representing one particular system (e.g. 1 = clear solution, 2 = emulsion, 3 = transparent gel etc.) and enclosing each system within a boundary, a phase diagram can be constructed. Suitable formulations, which will be clear solutions, become immediately apparent and the best can then be chosen, bearing in mind the properties required for this type of product.

It is also important to ensure that the formulation chosen does not lie too close to a phase boundary, as the positions of these can depend on the storage temperature of the product. In general the degree of solubilization of a drug increases as the temperature increases. From this type of phase diagram the physical composition of diluted preparations can also be shown. Point B, for example, represents a product consisting of 40% solubilizate and 60% surfactant. The construction of a straight line from here to point C represents the dilution of the product with increasing concentrations of water.

Should the concentration of drug to be included in the product be fixed, then the third axis can be used to represent varying concentrations of a third excipient such as a cosolvent. These values must, however, be plotted as percentage drug plus excipient to ensure a maximum value of 100%.

Alternatives to the use of surface-active agents as solubilizing agents include the cyclodextrins (Szejtli 1994). This range of compounds is based on a series of glucopyranose units that form cyclical structures resembling hollow cylinders (Fig. 21.3). As the inside surface of the ring is hydrophobic, owing to the presence of $-CH_2$ groups, drugs that are poorly soluble in water can be accommodated here. The outer part of the structure is hydrophilic and therefore freely soluble in water (Stella and Rajewski 1997).

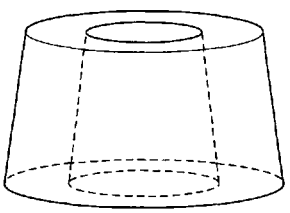


Fig. 21.3 Representation of the structure of cyclodextrins.

There are three natural cyclodextrins, the α, β and γ forms, the ring structures of which are composed of 6, 7 and 8 glucopyranose units, respectively, as well as an expanding series of derivatives.

Poorly soluble drugs of appropriate size slot into the interior of these structures, forming soluble inclusion complexes, usually with one 'host' molecule per cyclodextrin molecule (Pagington 1987). Examples of marketed drugs formulated with cyclodextrins include a range of prostaglandins for intra-arterial infusion, an antiseptic gargle based on iodine, and an oral liquid containing the antifungal drug itraconazole.

Complexation In some cases it may be possible to interact a poorly soluble drug with a soluble material to form a soluble intermolecular complex. As most complexes are macromolecular, however, they tend to be inactive, being unable to cross lipid membranes. It is therefore essential that complex formation is easily reversible, so that the free drug is released during or before contact with biological fluids.

It is not easy to predict whether a given drug will complex with a particular compound to improve solubility. Many complexes are not water soluble and may, in fact, be better suited for the prolonged release of the drug. Several well known examples are in general use, however, and include the complexation of iodine with a 10–15% solution of polyvinylpyrrolidone to improve the aqueous solubility of the active agent. Similarly, the interaction of salicylates and benzoates with xanthines, such as theophylline or caffeine, or with carbazochrome is carried out for the same effect.

Chemical modification As a last resort, chemical modification of a drug may be necessary in order to produce a water-soluble derivative. Examples include the synthesis of the sodium phosphate salts of hydrocortisone, prednisolone and betamethasone. The water-soluble chloramphenicol sodium succinate has no antibacterial activity of its own but is suitable for parenteral administration as a solution in order to obtain high blood levels, after which it is converted back to the less soluble active base. There are many examples of poorly soluble acids and bases being converted to a salt form to increase water solubility.

Particle size control The size and shape of very small particles, if less than 1 μm diameter, can affect their solubilities. As particle size decreases solubility will increase, and molecular dispersions of drugs in solid/solid solutions can exhibit improved bioavailability owing to the increase in solubility of the dispersed drug. In practice, however, this phenomenon has little application in the formulation of solutions, but is of particular relevance in suspension formulation.

Non-aqueous solutions

If it is not possible to ensure complete solution of the ingredients at all storage temperatures, or if the drug is unstable in aqueous systems it may be necessary to use an alternative, non-aqueous solvent. The use of non-aqueous systems may also have other advantages. For example, the intramuscular injection of solutions of drugs in oils is often used for depot therapy, and some drugs are specifically synthesized to improve their oil solubilities. The propionate and benzoate esters of testosterone and estradiol, respectively, are good examples of this. The oily solution remains as a discrete entity within the muscle tissue, releasing the drug slowly into the surrounding tissue by partitioning. A similar aqueous solution would diffuse readily and, being miscible with tissue fluid, would cause the drug to be released quickly.

It is essential that, in choosing a suitable solvent, its toxicity, irritancy and sensitizing potential are taken into account, as well as its flammability, cost, stability and compatibility with other excipients. It will be obvious that there is a greater choice of solvents available for inclusion into products for external application than those for internal use, and that for parenteral products the choice is limited even further.

A far wider range of solvents, however, is available for use as part of the manufacture of pharmaceutical products. In these instances the solvent is removed before packaging and is therefore not present in the final product. Examples include acetone, light petroleum and chloroform, although the latter is also used as a flavour and preservative in some extemporaneously prepared formulations. The following is a classification of some of the more widely used non-aqueous solvents in pharmaceutical preparations.

Fixed oils of vegetable origin

These are non-volatile oils that consist mainly of fatty acid esters of glycerol. Almond oil, for example, which consists of glycerides mainly of oleic acid, is used as a solvent for oily phenol injections, water being unsuitable because of the caustic nature of aqueous phenol solutions. Of similar chemical composition is arachis oil, which is used as the solvent in Dimercaprol Injection. Olive oil, sesame oil, maize oil, cottonseed oil, soya oil and castor oil are all suitable for parenteral use, the latter also being used as the solvent in miconazole eye drops (Lee 1985) and in some formulations of triamcinolone ear drops.

Ethyl oleate, which is a useful solvent for both Ergocalciferol Injection and Testosterone Propionate

Injection, is less viscous than the oils described above and therefore more easily injected intramuscularly. Of similar viscosity is benzyl benzoate, which can be used as an alternative solvent for dimercaprol.

Some fixed oils are sufficiently tasteless and odourless to be suitable for oral use as solvents for such materials as vitamins A and D. Fractionated coconut oil is occasionally used as the solvent for some antibiotics, which would otherwise hydrolyse rapidly if formulated as aqueous systems.

Veterinary formulations may also contain these solvents, arachis oil, for example, being used for hexachlorophene in the treatment of fascioliasis in ruminants.

Oils tend to be unpleasant to use externally, however, unless presented as an emulsion. Arachis oil is one of the few examples and is used as the solvent in Methyl Salicylate Liniment.

Alcohols

Ethyl alcohol is the most widely used solvent in this class, particularly for external application, where its rapid evaporation after application to the skin imparts a cooling effect to such products as salicylic acid lotion. It is also particularly useful for the extraction of crude drugs, being more selective than water. At concentrations greater than 15% ethanol exhibits antimicrobial activity, but because of its toxicity it is used orally or parenterally only at low concentrations, usually as a cosolvent with water.

If required for external use then industrial methylated spirit (IMS), which is free from excise duty, is usually included rather than the more expensive ethanol. Because industrial methylated spirit contains 5% methyl alcohol as a denaturant it is rendered too toxic for internal use.

An alcohol possessing similar properties is isopropyl alcohol, which is used externally as a solvent for dicophane. Its main advantage is that it is less likely to be abused than ethanol and that denaturation is not necessary.

Polyhydric alcohols

Alcohols containing two hydroxyl groups per molecule are known as glycols, but because of their toxicity they are rarely used internally. One important exception to this is propylene glycol.

$$CH_3.CH(OH)CH_2OH$$

which is often used in conjunction with water or glycerol as a cosolvent. It is used, for example, in the formulation of Digoxin Injection, Phenobarbital Injection, and some formulations of Diazepam Injection, Co-trimoxazole Intravenous Infusion and as the diluent for both Chloramphenicol Ear Drops and some brands of hydrocortisone ear drops, and in many preparations for oral use.

The lower molecular weight polyethylene glycols (PEG) or macrogols have the general formula:

$$HOCH_2(CH_2CH_2O)_nCH_2OH$$

and are available in a range of viscosity grades. PEG 400, for example, is used as a solvent in clotrimazole topical solution. They are also widely used as cosolvents with alcohol or water, although their main use is in the formulation of water-miscible ointment bases. There are also other glycols available which, although rarely included in products for human use, can be used for extraction processes or as solvents in the formulation of veterinary and horticultural solutions. Examples include dipropylene glycol, diethylene glycol, ethylene glycol and their monoethyl ethers. Glycerol, an alcohol possessing three hydroxyl groups per molecule, is also widely used, particularly as a cosolvent with water for oral use. At higher concentrations it is used in, for example, Phenol and Glycerol Injection.

Dimethylsulphoxide

This is a highly polar compound and is thought to aid the penetration of drugs through the skin. Although used mainly as a solvent for veterinary drugs, it is used as a carrier for idoxuridine, an antiviral agent, for application to human skin.

Ethyl ether

This material is widely used for the extraction of crude drugs, but because of its own therapeutic activity it is not used for the preparation of formulations for internal use. It is, however, used as a cosolvent with alcohol in some collodions.

Liquid paraffin

The oily nature of this material makes it unpleasant to use externally, although it is often used as a solvent for the topical application of drugs in emulsion formulations. At one time light liquid paraffin was widely used as the base for oily nasal drops. (These are now rarely used because of the possibility of causing lipoidal pneumonia if they are inhaled into the lungs.) It has a minor use in veterinary formulation as a solvent in, for example, anthelminthic drenches containing carbon tetrachloride.

Miscellaneous solvents

Isopropyl myristate and isopropyl palmitate are used as solvents for external use, particularly in cosmetics, where their low viscosity and lack of greasiness make them pleasant to use. Dimethylformamide and dimethylacetamide have both been used as solvents in veterinary formulation, but their toxicities render them unsuitable for human use. Kerosene too is also limited in its application, being used mainly as a solvent for insecticides such as pyrethrum and piperonyl butoxide.

Xylene is present in some ear drops for human use to dissolve ear wax, and glycofurol is a useful solvent for parenteral products.

As with aqueous systems, it may be possible to improve the solubility of a drug in a particular vehicle by the addition of a cosolvent. For example, nitrocellulose is poorly soluble in both alcohol and ether but adequately soluble in a mixture of both. The formulation of Digoxin Injection, too, is best achieved by the inclusion of both ethyl alcohol and propylene glycol.

OTHER FORMULATION ADDITIVES

Buffers

These are materials which, when dissolved in a solvent, will enable the solution to resist any change in pH should an acid or an alkali be added. The choice of suitable buffer depends on the pH and buffering capacity required. It must be compatible with other excipients and have a low toxicity. Most pharmaceutically acceptable buffering systems are based on carbonates, citrates, gluconates, lactates, phosphates or tartrates. Borates can be used for external application, but not to mucous membranes or to abraded skin.

Although solutions of drugs that are themselves weak electrolytes will act as buffers, their buffering capacities are not usually sufficiently robust and should be enhanced by one of the systems described above.

As the pH of most body fluids is 7.4, products such as injections, eye drops and nasal drops should, in theory, be buffered at this value to avoid irritation. Many body fluids themselves, however, have a buffering capacity and, when formulating low-volume intravenous injections or eye drops, a wider pH range can be tolerated. This is potentially useful should a compromise be necessary when choosing a pH that is physiologically acceptable for a drug whose optimum stability, solubility and/or bioavailability may depend on different pHs. For further details on the use of buffers see Chapters 3 and 35.

Density modifiers

It is rarely necessary to control the density of solutions except when formulating spinal anaesthetics. Solutions of lower density than cerebrospinal fluid will tend to rise after injection and those of higher density will fall. Careful control both of the density of such injections and of the position of the patient on the operating table will enable precise control of the area to be anaesthetized. The terms used to describe the density of injections in relation to that of spinal fluid are isobaric, hypobaric and hyperbaric, meaning of equal, lower and higher density, respectively. The most widely used material for density modification is dextrose.

Isotonicity modifiers

Solutions for injection, for application to mucous membranes, and large-volume solutions for ophthalmic use must be made iso-osmotic with tissue fluid to avoid pain and irritation.

The most widely used isotonicity modifiers are dextrose and sodium chloride. Isotonicity adjustments can only be made after the addition of all other ingredients, because each ingredient will contribute to the overall osmotic pressure of a solution.

Viscosity enhancement

It may be difficult for aqueous-based topical solutions to remain in place on the skin or in the eyes for any significant time because of their low viscosities. To counteract this effect, low concentrations of gelling agents can be used to increase the apparent viscosity of the product. Examples include povidone, hydroxyethylcellulose and carbomer.

Preservatives

When choosing a suitable preservative it must be ensured that:

- adsorption of the preservative onto the container from the product does not occur, and
- its efficiency is not impaired by the pH of the solution or by interactions with other ingredients.

For example, many of the widely used parahydroxybenzoic acid esters can be adsorbed into the micelles of some non-ionic surfactants and, although their presence can be detected by chemical analysis, they are in fact unable to exert their

antimicrobial activities. It is only by full microbiological challenge testing that the efficiency of a preservative system can be properly assessed.

A more comprehensive discussion on the preservation of pharmaceuticals can be found in Chapter 42.

Reducing agents and antioxidants

The decomposition of pharmaceutical products by oxidation can be controlled by the addition of reducing agents such as sodium metabisulphite, or antioxidants such as butylated hydroxyanisole or butylated hydroxytoluene. For unit-dose parenteral products, such as injections of nicotinamide and ascorbic acid, it is possible to use Water for Injections free from dissolved air and to replace the air in the headspace by nitrogen or another inert gas.

Sweetening agents

Low molecular weight carbohydrates, and in particular sucrose, are traditionally the most widely used sweetening agents. Sucrose has the advantage of being colourless, very soluble in water, stable over a pH range of about 4–8 and, by increasing the viscosity of fluid preparations, will impart to them a pleasant texture in the mouth. It will mask the tastes of both salty and bitter drugs and has a soothing effect on the membranes of the throat. For this reason, despite its cariogenic properties, sucrose is particularly useful as a vehicle for antitussive preparations.

Polyhydric alcohols such as sorbitol, mannitol and, to a lesser extent glycerol, also possess sweetening power and can be included in preparations for diabetic use, where sucrose is undesirable. Other less widely used bulk sweeteners include maltilol, lactilol, isomalt, fructose and xylitol. Treacle, honey and liquorice are now very rarely used, having only a minor application in some extemporaneously prepared formulations.

Artificial sweeteners can be used in conjunction with sugars and alcohols to enhance the degree of sweetness, or on their own in formulations for patients who must restrict their sugar intake. They are also termed intense sweeteners because, weight for weight, they are hundreds and even thousands of times sweeter than sucrose and are therefore rarely required at a concentration greater than about 0.2%.

Only about six artificial sweeteners are permitted for oral use within the European Union, the most widely used being the sodium or calcium salts of saccharin (E954). Both exhibit high water solubility and are chemically and physically stable over a wide pH range. Less widely used are aspartame (E951), which is a compound of L-aspartic acid and L-phenylalanine, acesulfame potassium (E950), thaumatin (E957), sodium cyclamate (E952) and neohesperidine DC (E959). The main disadvantage of all artificial sweeteners is their tendency to impart a bitter or metallic aftertaste, and they are therefore often formulated with sugars to mask this.

Flavours and perfumes

The simple use of sweetening agents may not be sufficient to render palatable a product containing a drug with a particularly unpleasant taste. In many cases, therefore, a flavouring agent can be included. This is particularly useful in paediatric formulation to ensure patient compliance. The inclusion of flavours has the additional advantage of enabling the easy identification of liquid products.

Flavouring and perfuming agents can be obtained from either natural or synthetic sources. Natural products include fruit juices, aromatic oils such as peppermint and lemon, herbs and spices, and distilled fractions of these. They are available as concentrated extracts, alcoholic or aqueous solutions, syrups or spirits, and are particularly widely used in the manufacture of products for extemporaneous use. Artificial perfumes and flavours are of purely synthetic origin, often having no natural counterpart. They tend to be cheaper, more readily available, less variable in chemical composition and more stable than natural products. They are usually available as alcoholic or aqueous solutions or as powders.

The choice of a suitable flavour can only be made as a result of subjective assessment and, as consumer preferences vary considerably, this is not easy. Some guidance can, however, be given by reference to Table 21.1, which shows that certain flavours are particularly useful for the masking of one or more of the basic taste sensations of saltiness, bitterness, sweetness and sourness. These tastes are detected by sensory receptors on various areas of the tongue,

Table 21.1 Suitable masking flavours for various product tastes

Taste of product	Suitable masking flavour
Salty	Apricot, butterscotch, liquorice, peach, vanilla
Bitter	Anise, chocolate, mint, passion fruit, wild cherry
Sweet	Vanilla, fruits, berries
Sour	Citrus fruits, liquorice, raspberry

whereas the more subtle flavours are detected by the olfactory receptors.

In some cases there is a strong association between the use of a product and its flavour or perfume content. For example, products intended for the relief of indigestion are often mint flavoured. This is because for many years mint has been used in such products for its carminative effect, but even in products containing other active agents the odour and taste of mint are now firmly associated with antacid activity. Similarly, the odour of terpineol is often associated with antiseptic activity and, in a competitive market, it may therefore be unwise to alter these flavours or perfumes to any appreciable extent.

The fact that personal preferences for flavours and perfumes often vary with age can also aid the formulator. Children, in general, prefer fruity tastes and smells, whereas adults choose flowery odours and acid flavours. Other suitable materials for the masking of unpleasant tastes include menthol, peppermint oil and chloroform. In addition to having their own particular tastes and odours they also act as desensitizing agents by exerting a mild anaesthetic effect on the sensory taste receptors.

Flavour-enhancing agents such as citric acid for citrus fruits and glycine or monosodium glutamate for general use are now becoming more widely used.

Colours

Once a suitable flavour has been chosen, it is often useful to include a colour associated with that flavour in order to improve the attractiveness of the product. Another reason for the inclusion of colours is to enable easy product identification, particularly of poisonous materials, including weedkillers and mineralized methylated spirit, and, for example, to differentiate between the many types of antiseptic solution used in hospitals for the disinfection of skin, instruments, syringes etc.

The presence of a strongly coloured degradation product, which does not affect the use of the product, may occasionally be masked by the use of a suitable colour. It is however, essential to ensure that any colour chosen is acceptable in the country in which the product is to be sold. A colour that is acceptable in one country may not be acceptable in another, and as aspects of colour legislation can change quite frequently it is necessary to ensure that only the latest regulations are consulted. The legal departments of most dye manufacturers are usually willing to supply up-to-date information.

The proliferation of nomenclature that exists for most colours can also cause confusion. For example, the water-soluble dye amaranth is also known as Bordeaux S, Cl Food Red 9 and Cl Acid Red 27. It has been allocated the Colour Index Number 16185 by the Society of Dyers and Colourists and the American Association of Textile Chemists and Colorists. Under the USA Food, Drug and Cosmetics Act it is known as FD and C Red Number 2, and a directive of the Council of European Communities has allocated it the reference number E 123.

As with flavours and perfumes, there is a range of both natural and synthetic colours. The former, which tend to be more widely acceptable, can be classified into carotenoids, chlorophyll, anthocyanins, and a miscellaneous group which includes riboflavines, caramel and extracts of beetroot. They can, however, exhibit the usual problems associated with natural products, namely variations in availability and chemical composition, both of which may cause formulation difficulties.

Synthetic or 'coal tar' dyes tend to give bright colours and are generally more stable than natural materials. Most of those that are suitable for pharmaceutical use are the sodium salts of sulphonic acids, and therefore they may be incompatible with cationic drugs. Care must also be taken to ensure that any dye used is not adversely affected by pH or by ultraviolet radiation, or by the inclusion of oxidizing or reducing agents or surfactants.

TYPES OF PREPARATION

The terminology used for the titles of different forms of liquid preparation for both oral and topical use can sometimes be confusing, with overlap between definitions and more than one definition being appropriate for one particular product. Furthermore, definitions can vary between different official compendia and may be at variance with definitions used within the pharmaceutical industry. This section attempts to give an overview of the types of pharmaceutical solution available.

Liquids for cutaneous application

Lotions, liniments, paints and collodions

Lotions can be formulated as solutions, and are designed to be applied to the skin without friction. They may contain humectants, so that moisture is retained on the skin after application of the product, or alcohol, which evaporates quickly, imparting a cooling effect and leaving the skin dry.

Liniments, however, are intended for massage into the skin and can contain such ingredients as methyl salicylate or camphor as counterirritants.

Liquids for application to the skin or mucous membranes in small amounts are often termed paints, and are usually applied with a small brush. The solvent is normally alcohol, acetone or ether, which evaporates quickly leaving a film on the skin that contains the active agent. A viscosity modifier such as glycerol is often added to ensure prolonged contact with the skin.

Collodions are similar preparations which, after evaporation of the solvent, leave a tough, flexible film that will seal small cuts or hold a drug in intimate contact with the skin. The film former is usually pyroxylin (nitrocellulose) in an alcohol/ether or alcohol/acetone solvent blend. Often a plasticizer such as castor oil and an adherent such as colophony resin are included.

Ear preparations

Also known as otic or aural products, these are simple solutions of drugs in either water, glycerol, propylene glycol or alcohol/water mixtures for local use, and include antibiotics, antiseptics, cleansing solutions and wax softeners. They are applied to the external auditory canal as drops, sprays or washes.

Eye preparations

These are small-volume sterile liquids designed to be instilled on to the eyeball or within the conjunctival sac for a local effect.

Irrigations

Irrigations are sterile, large-volume aqueous-based solutions for the cleansing of body cavities and wounds. They should be made isotonic with tissue fluid.

Mouthwashes and gargles

Aqueous solutions for the prevention and treatment of mouth and throat infections can contain antiseptics, analgesics and/or astringents. They are usually diluted with warm water before use.

Nasal products

These are formulated as small-volume solutions in an aqueous vehicle, oils being no longer used for nasal administration. Because the buffering capacity of nasal mucus is low, formulation at a pH of 6.8 is necessary. Nasal drops should also be made isotonic with nasal secretions using sodium chloride, and viscosity can also be modified using cellulose derivatives if necessary. Active agents for administration by this route for local use include antibiotics, anti-inflammatories and decongestants. The nasal route is also of major importance for specific types of drugs, and a full description of this method of drug delivery can be found in Chapter 32.

Oral liquids

This is a general term used to describe a solution, suspension or emulsion in which the active ingredient is dissolved or dispersed in a suitable liquid vehicle.

Elixirs

The terms mixture and elixir are often confused, although an elixir is strictly a solution of a potent or nauseous drug. If the active agent is sensitive to moisture, it may be formulated as a flavoured powder or granulation by the pharmaceutical industry and then simply dissolved in water immediately prior to administration. The dosage is usually taken using a 5 mL medicine spoon, although smaller volumes can be given using a volumetric dropper.

Linctuses

A linctus is a viscous preparation, usually prescribed for the relief of cough. It normally consists of a simple solution of the active agent in a high concentration of sucrose, often with other sweetening agents. This type of product, which is also designed to be administered in multiples of 5 mL, should be sipped slowly and not be diluted beforehand. The syrup content has a demulcent action on the mucous membranes of the throat. For diabetic use the sucrose is usually replaced by sorbitol and/or synthetic sweeteners.

Mixtures and draughts

Mixtures are usually aqueous preparations that can be in the form of either a solution or a suspension. Most preparations of this type are manufactured on a small scale as required, and are allocated a shelf-life of a few weeks before dispensing. Doses are usually given in multiples of 5 mL using a metric medicine spoon. A draught is a mixture of which only one or two large doses of about 50 mL are given, although smaller doses are often necessary for children.

Parenteral products

Sterile solutions for injection or infusion into the body are also available.

Rectal preparations

Aqueous or oily solutions, as well as emulsions and suspensions, are available for the rectal administration of medicaments for cleansing, diagnostic or therapeutic reasons, and are termed enemas.

Intermediate products

Aromatic waters and spirits

There are many pharmaceutical solutions that are designed for use during the manufacture of other preparations and which are rarely administered themselves. Aromatic waters, for example, are aqueous solutions of volatile materials and are used mainly for their flavouring properties. Examples include peppermint water and anise water, which also have carminative properties, and chloroform water, which also acts as a preservative.

They are usually manufactured as concentrated waters and are then diluted, traditionally 1:40 in the final preparation.

Spirits are also alcoholic solutions but of volatile materials, which are mainly used as flavouring agents.

Extracts, infusions and tinctures

Infusions, extracts and tinctures are terms used for concentrated solutions of active principles from animal or vegetable sources. Infusions are prepared by extracting the drug using 25% alcohol, but without the application of heat. Traditionally these preparations are then diluted 1:10 in the final product. Extracts are similar products that are then concentrated by evaporation. Tinctures are alcoholic extracts of drugs but are relatively weak compared with extracts.

Syrups

Syrups are concentrated solutions of sucrose or other sugars to which medicaments or flavourings are often added. For example, Codeine Phosphate Syrup is used as a cough suppressant and Orange Syrup contains dried bitter orange peel as a flavouring agent. Although syrups are used in the manufacture of other preparations, such as mixtures or elixirs, they can also be administered as products in their own right, the high concentrations of sugars imparting a sweetening effect.

As syrups can contain up to 85% of sugars, they are capable of resisting bacterial growth by virtue of their osmotic effect. Syrups can contain lower concentrations of sugars but will often include sufficient of a polyhydric alcohol such as sorbitol, glycerol or propylene glycol in order to maintain a high osmotic gradient. In addition, by acting as cosolvents they will help to prevent crystallization and to maintain solubility of all ingredients.

It is possible, however, in a closed container, for surface dilution of a syrup to take place. This occurs as a result of solvent evaporation that condenses on the upper internal surfaces of the container and then flows back on to the surface of the product, thereby producing a diluted layer which provides an ideal medium for the growth of certain microorganisms. For this reason syrups often contain additional preservatives.

A further problem with the storage and use of syrups involves the crystallization of the sugar within the screw cap used to seal the containers, thereby preventing its release. This can be avoided by the addition of the polyhydric alcohols previously mentioned, or by the inclusion of invert syrup, which is a mixture of glucose and fructose.

STABILITY OF SOLUTIONS

Both the chemical and the physical stability of solutions in their intended containers are important. A solution must retain its initial clarity, colour, odour, taste and viscosity over its allocated shelf-life. Clarity can easily be assessed by visual examination or by a measurement of its optical density after agitation. Colour too may be assessed both visually and spectrophotometrically, and equipment suitable for the measurement of rheological properties of solutions has been covered earlier.

The stability of flavours and perfumes is perhaps more difficult to assess. Although chromatographic methods are used with varying success to quantify these properties, considerable reliance must be placed on the organoleptic powers of a panel of assessors, who must be screened to ensure that their powers of olfaction and gustation are sufficiently sensitive. If a suitable majority of the panel members is unable to detect a difference between a stored sample and a freshly prepared reference material, it may be assumed that the taste or odour of the sample has not changed significantly.

MANUFACTURE OF SOLUTIONS

For both small- and large-scale manufacture of solutions the only equipment necessary is suitable mixing vessels, a means of agitation and a filtration system to ensure clarity of the final solution. During manufacture, the solute is simply added to the solvent in a mixing vessel and stirring is continued until dissolution is complete. If the solute is more soluble at elevated temperatures it may be advantageous to apply heat to the vessel, particularly if the dissolution rate is normally slow. Care must be taken, however, should any volatile or thermolabile materials be present. Size reduction of solid materials to increase their surface areas should also speed up the process of solution.

Solutes present in low concentrations, particularly dyes, are often predissolved in a small volume of the solvent and then added to the bulk. Volatile materials such as flavours and perfumes are, where possible, added at the end of a process and after any cooling, to reduce loss by evaporation. Finally, it must be ensured that significant amounts of any of the materials are not irreversibly adsorbed on to the filtration medium used for final clarification. For a discussion of suitable packaging materials and containers for solutions see Chapter 36.

REFERENCES

Garcia Sagredo, F., Guzman, M., Molpereces, J., Aberturas, M.R. (1994) Pluronic copolymers – characteristics, properties and pharmaceutical applications. *Pharm. Tech. Europe*, **1**, 46–56, **2**, 38–44.

Lee, R.L.H. (1985) *Aust. J. of Hosp. Pharm.*, **15**, 233

Pagington, I.S. (1987) β-cyclodextrin: the success of molecular inclusion. *Chemistry in Britain*, **23**, 455.

Stella, V.J. and Rajewski, R.A. (1997) Cyclodextrins: their future in drug formulation and delivery. *Pharm. Res.*, **14**, 556–567.

Szejtli, J. (1994) Medicinal applications of cyclodextrins. *Med. Res. Rev.*, **14**, 353–386.

BIBLIOGRAPHY

Martindale, W. (1999) *The Extra Pharmacopoeia* Pharmaceutical Press, London.

22
Clarification

Andrew Twitchell

CHAPTER CONTENTS

Clarification is a term used to describe processes that involve the removal or separation of a solid from a fluid, or a fluid from another fluid. The term 'fluid' encompasses both liquids and gases. Clarification can be achieved using either filtration or centrifugation techniques, both of which are described in this chapter.

In pharmaceutical processing there are two main reasons for such processes:

1. To remove unwanted solid particles from either a liquid product or from air;
2. To collect the solid as the product itself (e.g. following crystallization).

FILTRATION

Types of filtration

Solid/fluid filtration

Solid/fluid filtration can be defined as the separation of an insoluble solid from a fluid by means of a porous medium that retains the solid but allows the fluid to pass. It is the most common type of filtration encountered during the manufacture of pharmaceutical products. Solid/fluid filtration may be further subdivided into two types, namely solid/liquid filtration and solid/gas filtration.

Solid/liquid filtration There are numerous applications of solid/liquid filtration in pharmaceutical processing, some of which are listed below:

- Improvement of the appearance of solutions, mouthwashes etc. to give them a 'sparkle' or 'brightness'; this is often referred to as 'clarifying' a product;
- Removal of potential irritants, e.g. from eye-drop preparations or solutions applied to mucous membranes;

- Recovery of desired solid material from a suspension or slurry, e.g. to obtain a drug or excipient after a crystallization process;
- Certain operations, such as the extraction of vegetable drugs with a solvent, may yield a turbid product with a small quantity of fine suspended colloidal matter; this can be removed by filtration;
- Sterilization of liquid or semisolid products where processes involving heat (such as autoclaving) are not appropriate;
- Detection of microorganisms present in liquids. This can be achieved by analysing a suitable filter on which the bacteria are retained. This method can also be used to assess the efficiency of preservatives.

Solid/gas filtration There are two main applications of solid/gas filtration in pharmaceutical processing. One of particular importance in manufacturing is the removal of suspended solid material from air in order to supply air of the required standard for either processing equipment or manufacturing areas. This includes the provision of air for equipment such as fluidized-bed processors (see Chapters 25 and 26), film-coating machinery (Chapter 28) and bottle-cleaning equipment, so that product appearance and quality are maintained. The use of suitable filters also enables the particulate contamination of air in manufacturing areas to be at an appropriate level for the product being manufactured; for example, air free from microorganisms can be supplied to areas where sterile products are being manufactured.

It is also often necessary to remove particulate matter generated during a manufacturing operation from the process air in order to prevent the material being vented to the atmosphere. Examples of this include filtering of exhaust air from fluidized-bed and coating processes.

Fluid/fluid filtration

Flavouring oils are sometimes added to liquid preparations in the form of a spirit, i.e. dissolved in alcohol. When these spirits are added to aqueous-based formulations some of the oil may come out of solution, giving the product a degree of turbidity. Removal of the oil droplets by passing them through an appropriate filter (a liquid/liquid filtration process) is used to produce the desired product appearance.

Compressed air is used in a number of pharmaceutical processes (e.g. film-coating spray guns, bottle-cleaning equipment and fluid energy mills; see Chapter 11). Before use the compressed air needs to be filtered to ensure that any entrained oil or water droplets are removed. This is an example of a liquid/gas filtration process.

Mechanisms of filtration

The mechanisms by which material may be retained by a filter medium (i.e. the surface on or in which material is deposited) are discussed below.

Straining/sieving

If the pores in the filter medium through which the fluid is flowing are smaller than the material that is required to be removed, the material will be retained. Filtration occurs on the surface of the filter in this case, and therefore the filter can be very thin. Filter media of this type are referred to as ***membrane filters***. Because filtration occurs on the surface there is a tendency for them to become blocked unless the filter is carefully designed (see later). On a small scale, filters using the straining mechanism are used where the contaminant level is low or small volumes need to be filtered.

Examples of the use of membrane filters include the removal of bacteria and fibres from parenteral preparations.

Impingement

As a flowing fluid approaches and passes an object, for example a filter fibre, the fluid flow pattern is disturbed, as shown in Figure 22.1. Suspended solids may, however, have sufficient momentum that they do not follow the fluid path but impinge on the filter fibre and are retained, owing to attractive forces between the particle and the fibre. Where the pores between filter fibres are larger than the material being removed some particles may follow the fluid

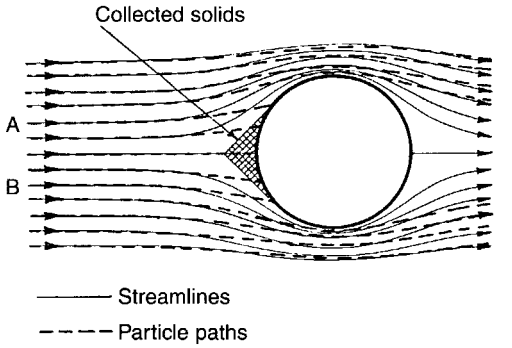


Fig. 22.1 Mechanisms of filtration by impingement.

streamlines and miss the fibre, this being more likely if the particles are small (owing to their lower momentum) and as the distance from the centre of the fibre towards which they approach increases. To ensure the removal of all unwanted material, filter media using the impingement mechanism must be sufficiently thick that material not trapped by the first fibre in its path is removed by a subsequent one. These types of filter are therefore referred to as depth filters. The fluid should flow through the filter medium in a streamlined manner to ensure the filter works effectively, as turbulent flow may carry the particles past the fibres. Depth filters are the main type of filter used for removing material from gases.

Attractive forces

Electrostatic and other surface forces may exert sufficient hold on the particles to attract and retain them on the filter medium (as occurs during the impingement mechanism).

Air can be freed from dust particles in an electrostatic precipitator by passing the air between highly charged surfaces, which attract the dust particles.

Autofiltration

Autofiltration is the term used to describe the situation when filtered material (termed the ***filter cake***) acts as its own filter medium. This mechanism is used by the metafilter which is covered later in this chapter.

Factors affecting the rate of filtration

The filtration process chosen must remove the required 'contaminants' or product but must also do so at an acceptably fast rate to ensure that the manufacturing process can be carried out economically. The laboratory Buchner funnel and flask (Fig. 22.2) is a convenient filter that can be used to illustrate the factors that influence the rate at which a product can be filtered. This filter is used for solid/liquid filtration processes, but the same basic principles are valid whatever filtration process is being evaluated.

The rate of filtration (volume of filtered material (V, m^3) obtained in unit time (t, s)) depends on the following factors:

1. The area available for filtration (A, m^2), which in this case is the cross-sectional area of the funnel;
2. The pressure difference (ΔP, Pa) across the filter bed (filter medium and any cake formed). With the Buchner funnel apparatus this difference is due to the 'head' of unfiltered product (slurry) and decreases as filtration proceeds and the level drops. Note that it is the pressure ***difference*** that is important, and this can be increased by drawing a vacuum in the collection flask. The difference between atmospheric pressure and the lower pressure in the flask is added to the pressure due to the unfiltered product to give the total pressure difference;

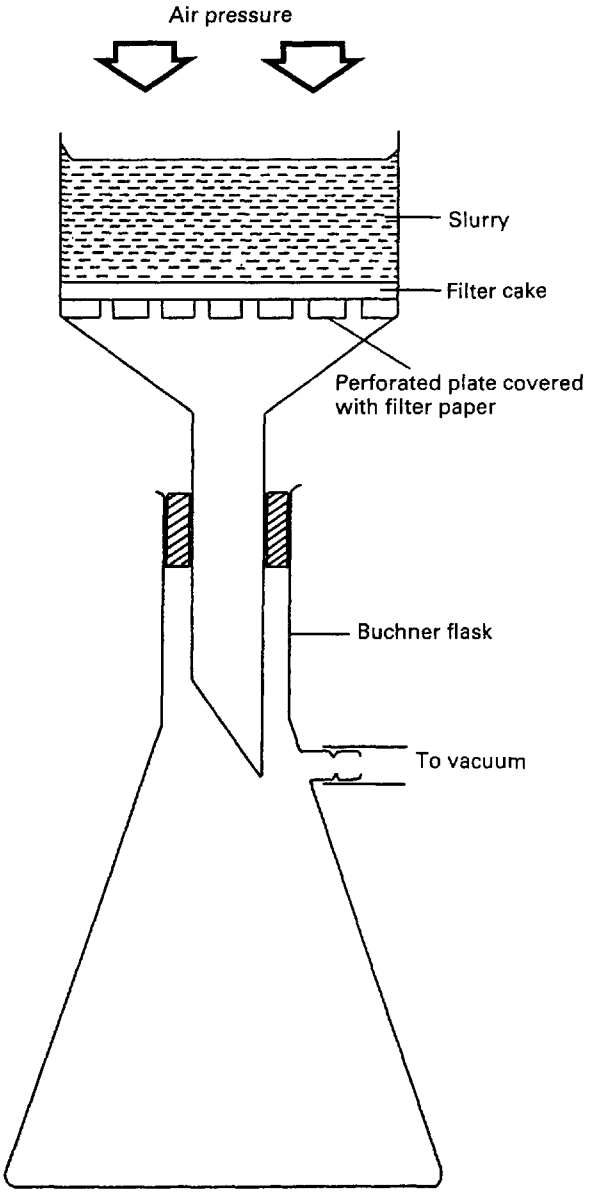


Fig. 22.2 Buchner funnel and vacuum flask.

3. The viscosity of the fluid passing through the filter, i.e. the filtrate (μ Pa s). A viscous fluid will filter more slowly than a mobile one owing to the greater resistance to movement offered by more viscous fluids (see Chapter 4).
4. The thickness of the filter medium and any deposited cake (L, m). The cake will **increase** in thickness as filtration proceeds, so if this is not removed the rate of filtration will fall.

The above factors are combined in the Darcy equation:

$$\frac{V}{t} = \frac{KA\Delta P}{\mu L} \qquad (22.1)$$

In this equation the driving force for this particular 'rate process' is the pressure difference across the filter, and the resistance to the process is a function of the properties of the filter bed, its thickness, and the viscosity of the filtrate. The contribution to resistance to filtration from the filter medium is usually small compared to that of the filter cake, and can often be neglected in calculations.

The proportionality constant K (m^2) expresses the ***permeability*** of the filter medium and cake and will increase as the porosity of the bed increases. It is clearly desirable that the K value should be large in order to maximize the filtration rate. If K is taken to represent the permeability of the cake it can be shown that K is given by:

$$K = \frac{e^3}{5(1-e)^2 S^2} \qquad (22.2)$$

where e is the porosity of the cake and S is the surface area of the particles comprising the cake. If the solid material is one that forms an impermeable cake the filtration rate may be improved by adding a **filter aid** (see later), which aids the formation of open porous cakes.

Methods used to increase filtration rate

Darcy's equation can be used to determine ways in which the filtration rate can be increased or controlled in practice. These are discussed below.

Increase the area available for filtration The total volume of filtrate flowing through the filter will be directly proportional to the area of the filter, and hence the rate of filtration can be increased by using either larger filters or a number of small units in parallel. Both of these approaches will also distribute the cake over a larger area and thus decrease the value of L, thereby further increasing the rate.

Increase the pressure difference across the filter cake The simplest filters, e.g. the laboratory filter funnel, use the gravitational force of the liquid 'head' to provide the driving force for filtration. Often this driving force is too low for a sufficiently quick filtration and there is a requirement to increase it. If a vacuum is 'pulled' on the far side of the filter medium (see Fig. 22.1) then the pressure difference can be increased up to atmospheric pressure, i.e. approximately 1×10^5 Pa, or 1 bar. In practice, however, it will be less, as the liquids will boil in the collecting vessel if the pressure is reduced to too low a value (see Chapter 39). Despite the limited pressure difference generated, vacuum filtration is used in the laboratory where there are safety advantages when using glassware, because if the glassware is damaged it will implode rather than explode. One important industrial filter – the rotary vacuum filter – also utilizes a vacuum; this is described later in this chapter.

With industrial-scale liquid filtration, commonly used means of obtaining a high-pressure difference are either pumping the material to be filtered into the filter using a suitable pump, or using a pressurized vessel to drive the liquid through the filter. Most industrial filters have positive-pressure feed, the pressure used being limited only by the pump and the ability of the filter to withstand the high-pressure stress. Pressures up to 15×10^5 Pa (15 bar) are commonly used.

Although increasing the ΔP value in the absence of any other changes will cause a proportional increase in filtration rate, care needs to be taken to ensure that a phenomenon known as **cake compression** does not occur. Too high an applied pressure may cause the particles making up the cake to deform and therefore decrease the voidage (bed porosity). It can be seen from Eqn 22.2 that small decreases in the value of the porosity (e) lead to large decreases in cake permeability (K), and therefore in the filtration rate. The effect on decreasing K greatly outweighs any increase in filtration rate arising from a thinner cake. There is also a danger of 'blinding' the filter medium at high pressures by forcing particles into it. This is most likely in the early stages before a continuous layer of cake has formed. As a general rule filtration should start at moderate pressure, which can be increased as filtration proceeds and the cake thickness builds up.

Decrease the filtrate viscosity The flow through a filter cake can be considered as the total flow through a large number of capillaries formed by the voids between the particles of the cake. The rate of flow through each capillary is governed by Poiseuille's law, which is a mathematical relationship that includes viscosity as a factor contributing to the

resistance to flow. To increase the filtration rate the viscosity of the filtrate can be reduced in most cases by heating the formulation to be filtered. Many industrial filters, e.g. the metafilter, can be fitted with a steam jacket which can control temperature and hence viscosity. Care needs to be taken with this approach, however, when filtering formulations containing volatile components, or if components are thermolabile. In such cases dilution of the formulation with water may be an alternative means of reducing the viscosity providing that the increase in filtration rate exceeds the effect of increasing the total volume to be filtered.

Decrease the thickness of filter cake Darcy's equation (Eqn 22.1) shows that the filtration rate falls off as the cake increases in thickness. This effect is commonly observed when filtering in the laboratory using filter paper in a funnel. In some cases if the cake is allowed to build up the process slows to an unacceptable rate, or may almost stop altogether. In these situations it may be necessary to remove the cake periodically or maintain it at a constant thickness, as occurs for example with the rotary drum filter. As previously mentioned, the cake thickness can be kept lower by using a large filter area.

Increase the permeability of the cake One way of increasing the permeability of the cake is to include filter aids. A filter aid is a material that, when included in the formulation to be filtered, forms a cake of a more open porous nature and thus increases the K value in Darcy's equation. In addition, it may reduce the compressibility of the cake and/or prevent the filtered material blocking the filter medium. Filter aids that are used include diatomite (a form of diatomaceous earth) and perlite, which is a type of volcanic glass. The use of filter aids is obviously not appropriate if the filtered material is the intended end product.

FILTRATION EQUIPMENT

The filtration equipment described in this chapter is that used for filtering liquids. Equipment for filtering gases (mainly air) are also available.

Equipment selection

Ideally the equipment chosen should allow a fast filtration rate to minimize production costs, be cheap to buy and run, be easily cleaned and resistant to corrosion, and be capable of filtering large volumes of product before the filter needs stripping down for cleaning or replacing.

There are a number of product-related factors that should be considered when selecting a filter for a particular process. These include:

- the chemical nature of the product. Interactions with the filter medium may lead to leaching of the filter components, degradation or swelling of the filter medium or adsorption of components of the filtered product on the filter. All of these may influence the efficiency of the filtration process or the quality of the filtered product;
- the volume to be filtered and the filtration rate required. These dictate the size and type of equipment and the amount of time needed for the filtration process;
- the operating pressure needed. This is important in governing the filtration rate (Eqn 22.1) and influences whether a vacuum filter (where the pressure difference is limited to 1×10^5 Pa) is appropriate. High operating pressures require that the equipment be of sufficient strength and that appropriate safe operating procedures be adopted;
- the amount of material to be removed. This will influence the choice of filter, as a large 'load' may necessitate the use of prefilters or may require a filter where the cake can be continuously removed;
- the degree of filtration required. This will dictate the pore size of membrane filters or the filter grade to be used. If sterility is required then the equipment should itself be capable of being sterilized, and must ensure that contamination does not occur after the product has passed the filter;
- the product viscosity and filtration temperature. A high product viscosity may require elevated pressures to be used. The incoming formulation can be heated, or steam-heated jackets be fitted to the equipment. Care should be taken to ensure the equipment seals etc. can operate at elevated temperatures.

Industrial filtration equipment

Filters for liquid products may be classified by the method used to drive the filtrate through the filter medium. Filters can be organized into three classes, namely gravity, vacuum and pressure filters.

Gravity filters

Filters that rely solely on gravity only generate low operating pressures, and therefore use on a large scale is limited. Gravity filters are, however, simple

and cheap, and are frequently used in laboratory filtration, where volumes are small and a low filtration rate is relatively unimportant.

Vacuum filters

The rotary vacuum filter In large-scale filtration continuous operation is often desirable, and this may be difficult when it is necessary to filter slurries containing a high proportion of solids. The rotary vacuum filter is continuous in operation and has a system for removing the cake so that it can be run for long periods handling concentrated slurries. A rotary drum filter is shown in section in Figure 22.3. It can be visualized as two concentric cylinders with the annular space between them divided into a number of septa by radial partitions. The outer cylinder is perforated and covered with a filter cloth. Each septum has a radial connection to a complicated rotating valve, whose function is to perform the sequence of operations listed in Table 22.1.

The cylinder rotates slowly in the slurry, which is kept agitated, and a vacuum applied to the segments draws filtrate into the septa, depositing cake on the filter cloth. When the deposited cake leaves the slurry bath vacuum is maintained to draw air through the cake, thus aiding drainage. This is followed by washing and then further drainage in the drying zone. The cake is removed by the scraper blade, aided by compressed air forced into the septa. It is the function of the rotary valve to direct these services into the septa where they are required.

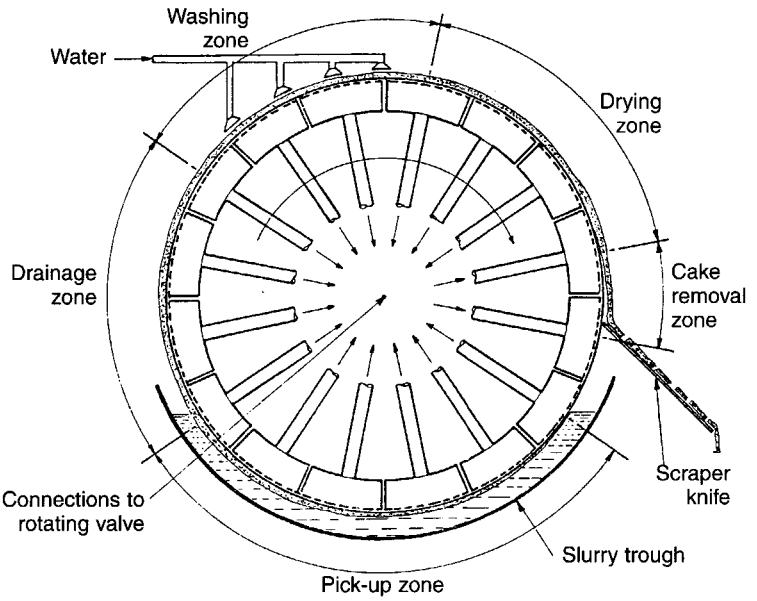


Fig. 22.3 Rotary drum filter.

Table 22.1 Rotary vacuum filter operation

Zone	Position	Service	Connected to
Pick-up	Slurry trough	Vacuum	Filtrate receiver
Drainage	–	Vacuum	Filtrate receiver
Washing	Wash sprays	Vacuum	Wash water receiver
Drying	–	Vacuum	Wash receiver
Cake removal	Scraper knife	Compressed air	Filter cake conveyor

In some cases, for example when the solid is the required product, the same receiver may he used for filtrate and for wash water.

Rotary filters can be up to 2 m in diameter and 3.5 m in length, with a filtration area of around 20 m^2. Cake compression rollers are often fitted to improve the efficiency of washing and draining if the cake on the drum becomes cracked. Difficult solids, which tend to block the filter cloth, necessitate a preliminary precoat of a thickness of filter aid to be deposited on the cloth prior to filtration of the slurry. During the actual filtration the scraper knife is set to move slowly inwards, removing the blocked outer layer of the filter aid and exposing fresh surface.

If removal of the cake presents problems, a string discharge filter may be employed. This is useful for filtration of the fermentation liquor in the manufacture of antibiotics, when a felt-like cake of mould mycelia must be removed. The filter cloth in this case has a number of loops of string passing round the drum and over two additional small rollers, as shown in Figure 22.4. In operation the strings lift the cake off the filter medium. The cake is broken by the sharp bend over the rollers and collected, and the strings return to the drum.

The advantages of the rotary vacuum filter can be summarized as follows:

1. It is automatic and continuous in operation, so that labour costs are very low.
2. The filter has a large capacity.
3. Variation of the speed of rotation enables the cake thickness to be controlled, and for solids that form an impenetrable cake the thickness may be limited to less than 5 mm. On the other hand, if the solids are coarse, forming a porous cake, the thickness may be 100 mm or more.

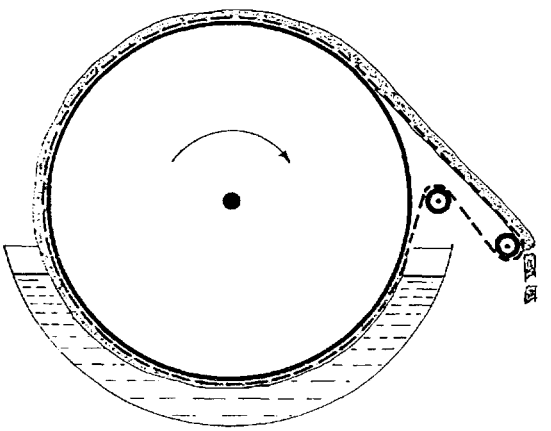


Fig. 22.4 String discharge rotary drum filter.

Disadvantages include:

1. The rotary filter is a complex piece of equipment with many moving parts and is very expensive. In addition to the filter itself, ancillary equipment such as vacuum pumps, vacuum receivers and traps, slurry pumps and agitators are required.
2. The cake tends to crack because of the air drawn through by the vacuum system, so that washing and drying are not efficient.
3. Being a vacuum filter, the pressure difference is limited to 1 bar and hot filtrates may boil.
4. The rotary filter is suitable only for straightforward slurries, being less satisfactory if the solids form an impermeable cake or will not separate cleanly from the cloth.

Uses of the rotary filter The rotary filter is most suitable for continuous operation on large quantities of slurry, especially if the slurry contains considerable amounts of solids, that is, in the range 15–30%.

Examples of pharmaceutical applications include the collection of calcium carbonate, magnesium carbonate and starch, and the separation of the mycelia from the fermentation liquor in the manufacture of antibiotics.

Pressure filters

Pressure filters feed the product to the filter at a pressure greater than that which would arise from gravity alone. This is the most common type of filter used in the processing of pharmaceutical products.

The metafilter In its simplest form, the metafilter consists of a grooved drainage rod on which is packed a series of metal rings. These rings, usually of stainless steel, are about 15 mm inside diameter, 22 mm outside diameter and 0.8 mm in thickness, with a number of semicircular projections on one surface (Fig. 22.5). The height of the projections and the shape of the section of the ring are such that when the rings are packed together, all the same way up, and tightened on the drainage rod with a nut, channels are formed that taper from about 250 μm down to 25 μm. One or more of these packs is mounted in a vessel, and the filter operated by pumping in the slurry under pressure.

In this form the metafilter can be used as a strainer for coarse particles, but for finer particles a bed of a suitable material (such as a filter aid) is first built up over the rings. The pack of rings, therefore, serves essentially as a base on which the true filter medium is supported.

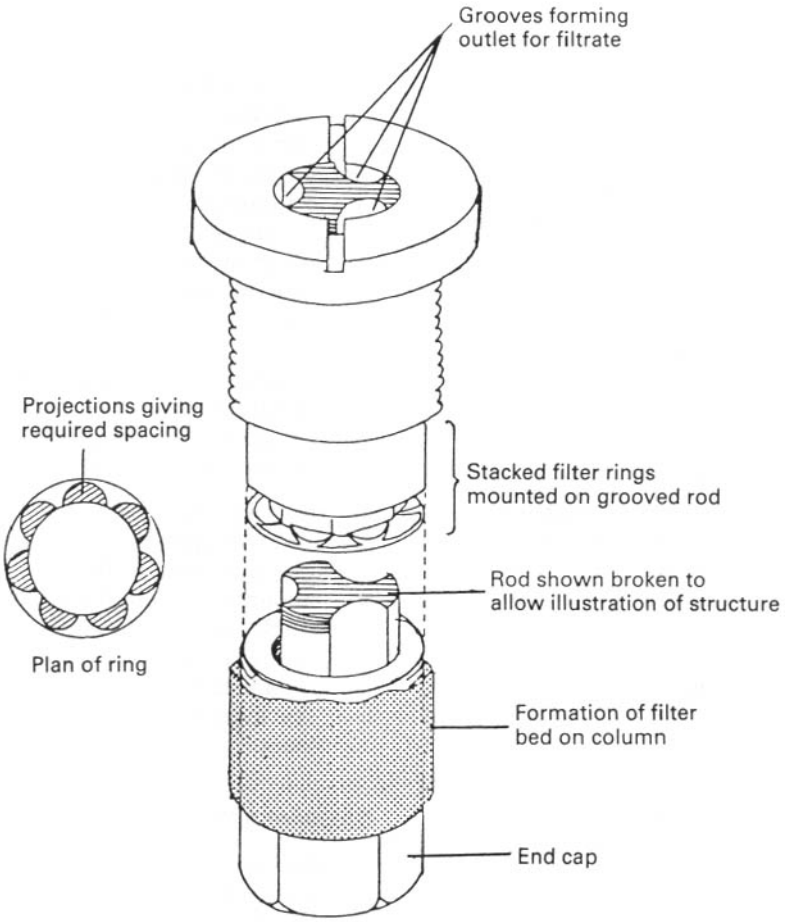


Fig. 22.5 Metafilter. Construction of the filter element (courtesy of Metafiltration Co. Ltd.).

The advantages of the metafilter can be summarized as follows:

1. It possesses considerable strength and high pressures can be used with no danger of bursting the filter medium.
2. As there is no filter medium as such, the running costs are low and it is very economical.
3. The metafilter can be made from materials (such as stainless steel) that can provide excellent resistance to corrosion and avoid contamination of the product.
4. By selecting of a suitable grade of material to form the filter bed it is possible to remove very fine particles; in fact, it is possible to sterilize a liquid using this filter

Uses of the metafilter The small surface area of the metafilter restricts the amount of solid that can be collected. This, together with the ability to separate very fine particles, means that the metafilter is used almost exclusively for clarifying liquids where the contaminant level is low.

Furthermore, the strength of the metafilter permits the use of high pressures (up to 15 bar), making the method suitable for viscous liquids.

Specific examples of pharmaceutical uses include the clarification of syrups, injection solutions, and products such as insulin liquors.

Cartridge filters Cartridge filters are now commonly used in the preparation of pharmaceutical products, as they possess a very large filtration area in a small unit and are easy and relatively cheap to operate. In simple form they consist of a cylindrical cartridge containing highly pleated material (e.g. PTFE or nylon) or 'string-wound' material (i.e. wound like a ball of string). This cartridge then fits in a metal supporting cylinder and the product is pumped under pressure into one end of the cylinder

surrounding the filter cartridge. The filtrate is forced through the filter cartridge from the periphery to the inner hollow core, from where it exits through the other end of the support cylinder. The filter cartridges are often disposable and are good for applications where there is a low contaminant level, e.g. during the filtration of liquid products as they are filled into bottles.

Cross-flow microfiltration It is possible to form membrane filters within 'hollow fibres'. The membrane, which may consist of polysulphone, acrylonitrile or polyamide, is laid down within a fibre which forms a rigid porous outer support (Fig. 22.6). The lumen of each fibre is small – typically 1–2 μm – but a large number of them can be contained in a surrounding shell to form a cartridge which may have an effective filtration area of over 2 m^2.

In use the liquid to be treated is pumped through the cartridge in a circulatory system, so that it passes through many times. The filtrate, which in this technique is often called the 'permeate', flows ***radially*** through the membrane and porous support. The great advantage of this mode of operation is that the high fluid velocity and turbulence minimize blocking of the membranes. Fresh liquid enters the system from a reservoir as filtration proceeds. Because the fluid flows ***across*** the surface, rather than at right-angles, this technique is known as ***cross-flow*** microfiltration.

Uses The method has been used for fractionation of biological products by first using a filter of pore size sufficient to let all the wanted molecules through, and then passing the permeate through a filter which will retain the required molecules while passing smaller unwanted molecules. It is claimed that blood plasma can be processed to remove alcohol and water and prepare concentrated purified albumin. The process has been suggested for the recovery of antibiotics from fermentation media.

CENTRIFUGATION

Centrifugal force can be used either to provide the driving force (ΔP) for the filtration process (refer to Darcy's law above, Eqn 22.1) or to replace the gravitational force in sedimentation processes (refer to Stokes' law, Chapter 4). Centrifuges are often used in the laboratory to separate solid material from a liquid, the solid typically forming a 'plug' at the bottom of the test tube at the end of the process.

Principles of centrifugation

If a particle (mass = m kg) spins in a centrifuge (radius r m) at a velocity (v m s^{-1}) then the centrifugal force (F N) acting on the particle equals $m\ v^2/r$.

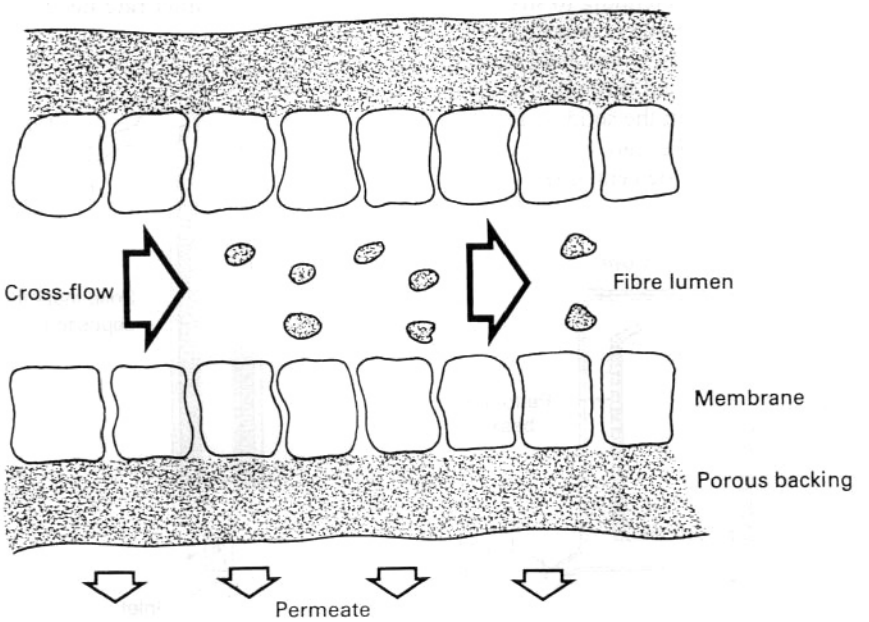


Fig. 22.6 Cross-flow microfiltration through an individual fibre

The same particle experiences gravitational force (G, N) = $m\,g$ (where g = gravitational constant).

The ***centrifugal effect*** (C) is the ratio of these two forces, so $C = F/G$, i.e. C indicates how much greater F is than G. Therefore, $C = v^2/g\,r$. If the velocity is taken to be $\pi\,d\,n$, where n is rotation speed (s^{-1}) and d is diameter of rotation (m), then $C = 2.01\,d\,n^2$.

In order to increase the centrifugal effect it is therefore more efficient to increase the centrifuge speed than to use a larger diameter at the same speed.

Larger centrifuges also generate greater pressures on the centrifuge wall for the same value of C, and are therefore more costly to make.

Industrial centrifuges

There are two main types of centrifuge used to achieve separation on an industrial scale, those using perforated baskets, which perform a filtration-type operation (work like a spin-dryer) and those with a solid walled vessel, where particles sediment towards the wall under the influence of the centrifugal force.

Perforated-basket centrifuges (centrifugal filters)

A diagram of a perforated basket centrifuge is shown in Figure 22.7. It consists of a stainless steel perforated basket (typically 1–2 m in diameter) lined with a filter cloth. The basket rotates at a speed which is typically <25 s^{-1}, higher speeds tending to stress the basket excessively. The product enters centrally and is thrown outwards by centrifugal force and held against the filter cloth. The filtrate is forced through the cloth and removed via the liquid outlet; the solid material is retained on the cloth. The cake can be washed if required by spraying water into the centrifuge.

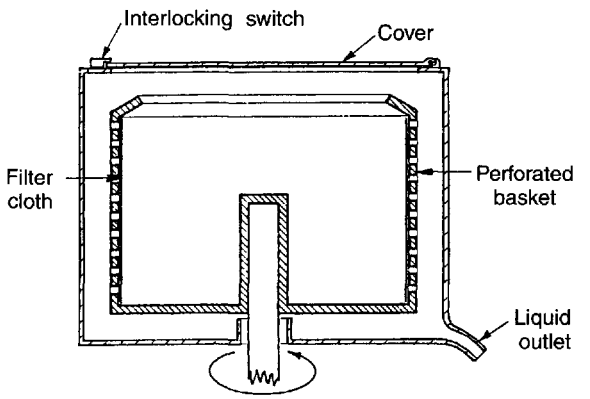


Fig. 22.7 Centrifugal dewatering filter.

The centrifugal filter has been used for separating crystalline materials from the preparation liquor, e.g. in the preparation of aspirin, and for removing precipitated proteins from insulin. It has the advantages of being compact and efficient, a 1 m centrifuge being able to process about 200 kg in 10 minutes. It can also handle concentrated slurries which might block other filters, and gives a product with a very low moisture content (typically around 2% w/w), which saves energy during drying.

The centrifuge described above is operated batchwise but continuous centrifuges are available for large-scale work. These have means for automatic discharge of the cake from a basket, which rotates around a horizontal axis in contrast to the vertical axis. Most of the energy required to run a centrifuge is used to bring it up to operating speed and little more is needed to maintain that speed. Continuous centrifuges are therefore cheaper to run, but the initial cost is considerably more.

Tubular-bowl centrifuges (centrifugal sedimenters)

These consist of a cylindrical 'bowl', typically around 100 mm in diameter and 1 m long, which rotates at 300–1000 s^{-1}. The product enters at the bottom and centrifugal force causes solids to be deposited on the wall as it passes up the bowl, the liquid overflowing from the top (Fig. 22.8). This type of centrifuge can also be adapted to separate immiscible liquids. The inlet rate needs to be controlled so

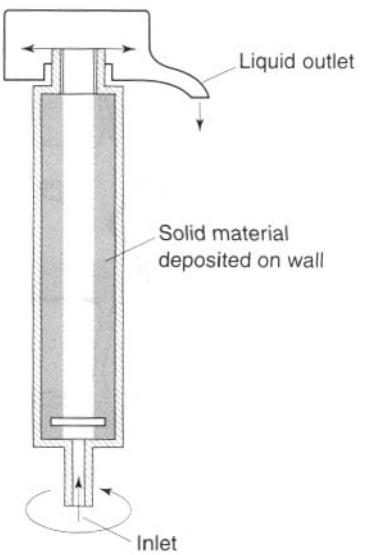


Fig. 22.8 Tubular-bowl centrifuge.

that there is sufficient time for sedimentation to occur before the product leaves the bowl.

The uses of centrifugal sedimenters include liquid/liquid separation, e.g. during antibiotic manufacture and purification of fish oils, the removal of very small particles, the removal of solids that are compressible or 'slimy' and which easily block the filter medium, the separation of blood plasma from whole blood (need $C \approx 3000$), the separation of different particle size fractions, and examining the stability of emulsions.

These centrifuges are compact, have a high separating efficiency and are good for separating 'difficult' solids, but have a limited capacity and are complicated to construct to achieve the required speed and minimize vibration.

Bibliography

Meltser, T.H. (1987) *Filtration in the Pharmaceutical Industry*, Marcel Dekker.

23
Suspensions and emulsions

Michael Billany

CHAPTER CONTENTS

INTRODUCTION

A coarse suspension is a dispersion of finely divided, insoluble solid particles (the disperse phase) in a fluid (the dispersion medium or continuous phase). Most pharmaceutical suspensions consist of an aqueous dispersion medium, although in some instances it may be an organic or oily liquid. A disperse phase with a mean particle diameter of up to 1 μm is usually termed a colloidal dispersion, and includes such examples as aluminium hydroxide and magnesium hydroxide suspensions. A solid in liquid dispersion, in which the particles are above colloidal size, is termed a coarse suspension.

An emulsion may be defined as two immiscible liquids, one of which is finely subdivided and uniformly distributed as droplets throughout the other. The system is stabilized by the presence of an emulsifying agent. The dispersed liquid or internal phase usually consists of globules of diameters down to 0.1 μm which are distributed within the external or continuous phase.

The physical properties of both colloidal and coarse suspensions and of emulsions are discussed in Chapter 6.

PHYSICAL PROPERTIES OF WELL-FORMULATED SUSPENSIONS AND EMULSIONS

- The product must remain sufficiently homogenous for at least the period between shaking the container and removing the required amount.
- The sediment or creaming produced on storage, if any, must be easily resuspended by moderate agitation of the container.
- The product may be required to be thickened in order to reduce the rate of settling of the particles or the rate of creaming of oil globules. The resulting viscosity must not be so high that removal of the product from the container and transfer to the site of application are difficult.
- Any suspended particles should be small and uniformly sized in order to give a smooth, elegant product, free from a gritty texture.

PHARMACEUTICAL APPLICATIONS OF SUSPENSIONS

Suspensions can be used as oral dosage forms, applied topically to the skin or mucous membrane surfaces, or given parenterally by injection.

Suspensions as oral drug delivery systems

Many people have difficulty in swallowing solid dosage forms and therefore require the drug to be dispersed in a liquid.

Some materials are required to be present in the gastrointestinal tract in a finely divided form, and their formulation as suspensions will provide the desired high surface area. Solids such as kaolin, magnesium carbonate and magnesium trisilicate, for example, are used for the adsorption of toxins, or to neutralize excess acidity. A dispersion of finely divided silica in dimethicone 1000 is used in veterinary practice for the treatment of 'frothy bloat'.

The taste of most drugs is more noticeable if it is in solution rather than in an insoluble form. Paracetamol is available both in solution as Paediatric Paracetamol Oral Solution and also as a suspension. The latter is more palatable, and therefore particularly suitable for children. For the same reason chloramphenicol mixtures can be formulated as suspensions containing the insoluble chloramphenicol palmitate.

Suspensions for topical administration

Suspensions of drugs can also be formulated for topical application (Chapter 33). They can be fluid preparations, such as Calamine Lotion, which are designed to leave a light deposit of the active agent on the skin after quick evaporation of the dispersion medium. Some suspensions, such as pastes, are semisolid in consistency and contain high concentrations of powders dispersed – usually – in a paraffin base. It may also be possible to suspend a powdered drug in an emulsion base, as in Zinc Cream.

Suspensions for parenteral use and inhalation therapy

Suspensions can also be formulated for parenteral administration in order to control the rate of absorption of the drug. By varying the size of the dispersed particles of active agent, the duration of activity can be controlled. The absorption rate of the drug into the bloodstream will then depend simply on its rate of dissolution. If the drug is suspended in a fixed oil such as arachis or sesame, the product will remain after injection in the form of an oil globule, thereby presenting to the tissue fluid a small surface area from which the partitioning of drug can occur. The release of drug suspended in an aqueous vehicle will be faster, as some diffusion of the product will occur along muscle fibres and become miscible with tissue fluid. This will present a larger surface area from which the drug can be released.

Vaccines for the induction of immunity are often formulated as dispersions of killed microorganisms, as in Cholera Vaccine, or of the constituent toxoids adsorbed on to a substrate of aluminium hydroxide or phosphate, as in Adsorbed Diphtheria and Tetanus Vaccine. Thus a prolonged antigenic stimulus is provided, resulting in a high antibody titre.

Some X-ray contrast media are also formulated in this way. Barium sulphate, for the examination of the alimentary tract, is available as a suspension for either oral or rectal administration, and propyliodone is dispersed in either water or arachis oil for examination of the bronchial tract.

The adsorptive properties of fine powders are also used in the formulation of some inhalations. The volatile components of menthol and eucalyptus oil would be lost from solution very rapidly during use, whereas a more prolonged release is obtained if the two active agents are adsorbed on to light magnesium carbonate prior to the preparation of a suspension.

Chapter 31 describes some aspects of the formulation of aerosols, many of which are also available as suspensions of the active agent in a mixture of propellants.

Solubility and stability considerations

If the drug is insoluble or poorly soluble in a suitable solvent, then formulation as a suspension is usually required. Some eye drops, notably Hydrocortisone Acetate and Neomycin Eye Drops, are formulated as suspensions because of the poor solubility of hydrocortisone in a suitable solvent.

The degradation of a drug in the presence of water may also preclude its use as an aqueous solution. In this case it may be possible to synthesize an insoluble derivative that can then be formulated as a suspension. For example, oxytetracycline hydrochloride is used in solid dosage forms, but in aqueous solution would rapidly hydrolyse. A stable liquid dosage form has been made by suspending the insoluble calcium salt in a suitable aqueous vehicle.

Prolonged contact between the solid drug particles and the dispersion medium can be considerably reduced by preparing the suspension immediately prior to issue to the patient. Amoxicillin, for example, is provided by the manufacturer as the trihydrate salt mixed with the other powdered or granulated ingredients. The pharmacist then makes the product up to volume with water immediately before issue to the patient, allocating a shelf-life of 14 days at a temperature at or below 25°C.

A drug that degrades in the presence of water may alternatively be suspended in a non-aqueous vehicle. Fractionated coconut oil is used as the vehicle for some formulations of antibiotics for oral use, and in some countries tetracycline hydrochloride is dispersed in a similar base for ophthalmic use.

FORMULATION OF SUSPENSIONS

Particle size control

It is first necessary to ensure that the drug to be suspended is of a fine particle size prior to formulation. This is to ensure a slow rate of sedimentation of the suspended particles. Large particles, if greater than about 5 μm diameter, will also impart a gritty texture to the product, and may cause irritation if injected or instilled into the eyes. The ease of administration of a parenteral suspension may depend upon particle size and shape, and it is quite possible to block a hypodermic needle with particles over about 25 μm diameter, particularly if they are acicular in shape rather than isodiametric. A particular particle size range may also be chosen in order to control the rate of dissolution of the drug and hence its bioavailability.

Even though the particle size of a drug may be small when the suspension is first manufactured, there is always a degree of crystal growth that occurs on storage, particularly if temperature fluctuations occur. This is because the solubility of the drug may increase as the temperature rises, but on cooling, the drug will crystallize out. This is a particular problem with slightly soluble drugs such as paracetamol.

If the drug is polydispersed, then the very small crystals of less than 1 μm diameter will exhibit a greater solubility than the larger ones. Over a period of time the small crystals will become even smaller, whereas the diameters of the larger particles will increase. It is therefore advantageous to use a suspended drug of a narrow size range. The inclusion of surface-active agents or polymeric colloids, which adsorb on to the surface of each particle, may also help to prevent crystal growth.

Different polymorphic forms of a drug may exhibit different solubilities, the metastable state being the most soluble. Conversion of the metastable form, in solution, to the less soluble stable state, and its subsequent precipitation, will lead to changes in particle size.

The use of wetting agents

Some insoluble solids may be easily wetted by water and will disperse readily throughout the aqueous phase with only minimal agitation. Most, however, will exhibit varying degrees of hydrophobicity and will not be easily wetted. Some particles will form large porous clumps within the liquid, whereas others remain on the surface and become attached to the upper part of the container. The foam produced on shaking will be slow to subside because of the stabilizing effect of the small particles at the liquid/air interface.

To ensure adequate wetting, the interfacial tension between the solid and the liquid must be reduced so that the adsorbed air is displaced from the solid surfaces by the liquid. The particles will then disperse readily throughout the liquid, particularly if an intense shearing action is used during mixing. If a series of suspensions is prepared, each containing one of a range of concentrations of wetting agent, then the concentration to choose will be the lowest that provides adequate wetting.

The following is a discussion of the most widely used wetting agents for pharmaceutical products.

Surface-active agents

Figure 6.15 shows that surfactants possessing an HLB value between about 7 and 9 would be suitable for use as wetting agents. The hydrocarbon chains would be adsorbed by the hydrophobic particle surfaces, whereas the polar groups project into the aqueous medium and become hydrated. Wetting of the solid occurs as a result of a fall both in interfacial tension between the solid and the liquid and, to a lesser extent, between the liquid and air.

Most surfactants are used at concentrations of up to about 0.1% as wetting agents and include, for oral use, the polysorbates (Tweens) and sorbitan esters (Spans). For external application, sodium lauryl sulphate, sodium dioctylsulphosuccinate and quillaia extract can also be used.

The choice of surfactant for parenteral administration is obviously more limited, the main ones used

being the polysorbates, some of the poloxamers (polyoxyethylene/polyoxypropylene copolymers) and lecithin.

Disadvantages in the use of this type of wetting agent include excessive foaming and the possible formation of a deflocculated system, which may not be required.

Hydrophilic colloids

These materials include acacia, bentonite, tragacanth, alginates, xanthan gum and cellulose derivatives, and will behave as protective colloids by coating the solid hydrophobic particles with a multimolecular layer. This will impart a hydrophilic character to the solid and so promote wetting. These materials are also used as suspending agents and may, like surfactants, produce a deflocculated system, particularly if used at low concentrations.

Solvents

Materials such as alcohol, glycerol and glycols, which are water miscible, will reduce the liquid/air interfacial tension. The solvent will penetrate the loose agglomerates of powder displacing the air from the pores of the individual particles, so enabling wetting to occur by the dispersion medium.

Flocculated and deflocculated systems

Having incorporated a suitable wetting agent, it is then necessary to determine whether the suspension is flocculated or deflocculated and to decide which state is preferable. Whether or not a suspension is flocculated or deflocculated depends on the relative magnitudes of the forces of repulsion and attraction between the particles. The effects of these particle–particle interactions have been adequately covered in Chapter 6.

In a deflocculated system the dispersed particles remain as discrete units and, because the rate of sedimentation depends on the size of each unit, settling will be slow. The supernatant of a deflocculated system will continue to remain cloudy for an appreciable time after shaking, due to the very slow settling rate of the smallest particles in the product, even after the larger ones have sedimented. The repulsive forces between individual particles allow them to slip past each other as they sediment. The slow rate of settling prevents the entrapment of liquid within the sediment, which thus becomes compacted and can be very difficult to redisperse. This phenomenon is also called caking or claying, and is the most serious of all the physical stability problems encountered in suspension formulation.

The aggregation of particles in a flocculated system will lead to a much more rapid rate of sedimentation or subsidence because each unit is composed of many individual particles and is therefore larger. The rate of settling will also depend on the porosity of the aggregate, because if it is porous the dispersion medium can flow through, as well as around, each aggregate or floccule as it sediments.

The nature of the sediment of a flocculated system is also quite different from that of a deflocculated one. The structure of each aggregate is retained after sedimentation, thus entrapping a large amount of the liquid phase. As explained in Chapter 6, aggregation in the primary minimum will produce compact floccules, whereas a secondary minimum effect will produce loose floccules of higher porosity. Whichever occurs, the volume of the final sediment will still be large and will easily be redispersed by moderate agitation.

In a flocculated system the supernatant quickly becomes clear, as the large flocs that settle rapidly are composed of particles of all sizes. Figure 23.1 illus-

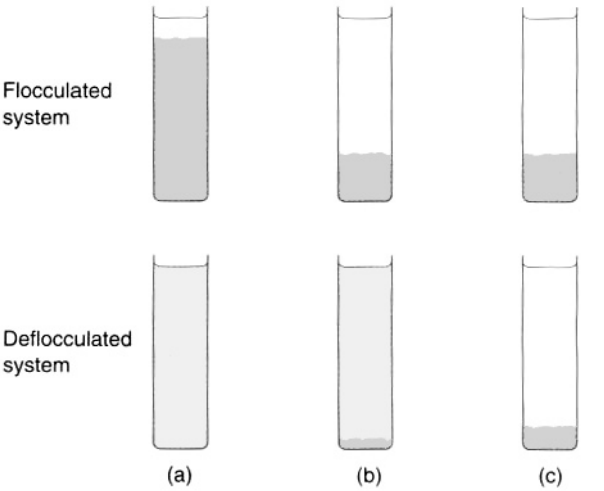


Fig. 23.1 The sedimentation behaviour of flocculated and deflocculated suspensions. Within a few minutes of manufacture (a) there is no apparent change within the deflocculated system compared to its initial appearance. Even after several hours (b) there is still little obvious change, except that the concentration of solids in the lower layers has increased at the expense of the upper layers owing to slow particle sedimentation. There is a small amount of a compact sediment. After prolonged storage (c), depending on the physical stability of the system, the supernatant has cleared, leaving a compact sediment. In the flocculated system at (a) there is some clear supernatant with a distinct boundary between it and the sediment. At (b) there is a larger volume of clear supernatant with a relatively large volume of a porous sediment, which does not change further even after prolonged storage (c).

trates the appearance of both flocculated and deflocculated suspensions at given times after shaking.

In summary, deflocculated systems have the advantage of a slow sedimentation rate, thereby enabling a uniform dose to be taken from the container, but when settling does occur the sediment is compacted and difficult to redisperse. Flocculated systems form loose sediments which are easily redispersible, but the sedimentation rate is fast and there is a danger of an inaccurate dose being administered; also, the product will look inelegant.

Controlled flocculation

A deflocculated system with a sufficiently high viscosity to prevent sedimentation would be an ideal formulation. It cannot be guaranteed, however, that the system would remain homogenous during the entire shelf-life of the product. Usually a compromise is reached in which the suspension is partially flocculated to enable adequate redispersion if necessary, and viscosity is controlled so that the sedimentation rate is at a minimum.

The next stage of the formulation process, after the addition of the wetting agent, is to ensure that the product exhibits the correct degree of flocculation. Underflocculation will give those undesirable properties that are associated with deflocculated systems. An overflocculated product will look inelegant and, to minimize settling, the viscosity of the product may have to be so high that any necessary redispersion would be difficult.

Controlled flocculation is usually achieved by a combination of particle size control, the use of electrolytes to control zeta potential, and the addition of polymers to enable crosslinking to occur between particles. Some polymers have the advantage of becoming ionized in an aqueous solution, and can therefore act both electrostatically and sterically. These materials are also termed polyelectrolytes.

Flocculating agents

In many cases, after the incorporation of a non-ionic wetting agent a suspension will be found to be deflocculated, either because of the reduction in solid/liquid interfacial tension, or because of the hydrated hydrophilic layer around each particle forming a mechanical barrier to aggregation. The use of an ionic surfactant to wet the solid could produce either a flocculated or a deflocculated system, depending on any charge already present on the particles. If particles are of opposite charge to that of the surfactant then neutralization will occur. If a high charge density is imparted to the suspended particles then deflocculation will be the result.

If it is necessary for the suspension to be converted from a deflocculated to a partially flocculated state, this may be achieved by the addition of electrolytes, surfactants and/or hydrophilic polymers.

Electrolytes The addition of an inorganic electrolyte to an aqueous suspension will alter the zeta potential of the dispersed particles and, if this value is lowered sufficiently, flocculation may occur.

The Schultz–Hardy rule shows that the ability of an electrolyte to flocculate hydrophobic particles depends on the valency of its counter-ions. Although they are more efficient, trivalent ions are less widely used than mono- or divalent electrolytes because they are generally more toxic. If hydrophilic polymers, which are usually negatively charged, are included in the formulation they may be precipitated by the presence of trivalent ions.

The most widely used electrolytes include the sodium salts of acetates, phosphates and citrates, and the concentration chosen will be that which produces the desired degree of flocculation. Care must be taken not to add excessive electrolyte or charge reversal may occur on each particle, so forming, once again, a deflocculated system.

Surfactants Ionic surface-active agents may also cause flocculation by neutralizing the charge on each particle, thus resulting in a deflocculated system. Non-ionic surfactants will, of course, have a negligible effect on the charge density of a particle but may, because of their linear configurations, adsorb on to more than one particle, thereby forming a loose flocculated structure.

Polymeric flocculating agents Starch, alginates, cellulose derivatives, tragacanth, carbomers and silicates are examples of polymers that can be used to control flocculation. Their linear branched-chain molecules form a gel-like network within the system and become adsorbed on to the surfaces of the dispersed particles, thus holding them in a flocculated state. Although some settling can occur, the sedimentation volume is large, and usually remains so for a considerable period.

Care must be taken to ensure that, during manufacture, blending is not excessive as this may inhibit the crosslinking between adjacent particles and result in the adsorption of each molecule of polymer on to one particle only. If this should occur then a deflocculated system may result, because the formation of the hydrophilic barrier around each particle will inhibit aggregation. A high concentration of polymer may have a similar effect if the whole surface of each particle is coated. It is essential that

areas on each suspended particle remain free from adsorbate, so that crosslinking can recur after the product is sheared. Further details of the use of polymers can be found in the next section.

Rheology of suspensions

An ideal pharmaceutical suspension would exhibit a high apparent viscosity at low rates of shear so that, on storage, the suspended particles would either settle very slowly or, preferably, remain permanently suspended. At higher rates of shear, such as those caused by moderate shaking of the product, the apparent viscosity should fall sufficiently for the product to be poured easily from its container. The product, if for external use, should then spread easily without excessive dragging, but should not be so fluid that it runs off the skin surface. If intended for injection, the product should pass easily through a hypodermic needle with only moderate pressure applied to the syringe plunger. It would then be important for the initial high apparent viscosity to be reformed after a short time to maintain adequate physical stability.

A flocculated system partly fulfils these criteria. In such a system pseudoplastic or plastic behaviour (see Chapter 4) is exhibited as the structure progressively breaks down under shear. The product then shows the time-dependent reversibility of this loss of structure, which is termed thixotropy.

A deflocculated system, however, would exhibit newtonian behaviour owing to the absence of such structures and may even, if high concentrations of disperse phase are present, exhibit dilatancy.

Although a flocculated system may exhibit some thixotropy and plasticity, unless a high concentration of disperse phase is present it may not be sufficient to prevent rapid settling, particularly if a surfactant or an electrolyte is present as a flocculating agent. In these cases suspending agents may be used to increase the apparent viscosity of the system.

Suitable materials are the hydrophilic polymers discussed above. These exert their effect by entrapping the solid dispersed particles within their gel-like network, so preventing sedimentation. At low concentrations many suspending agents can be used to control flocculation, and it must be realized that if large quantities are to be used to enhance viscosity the degree of flocculation may also be altered.

Viscosity modifiers

The following materials are those most widely used for the modification of suspension viscosity.

Polysaccharides

Acacia This natural material is often used as a suspending agent for extemporaneously prepared suspensions. Acacia is not a good thickening agent and its value as a suspending agent is largely due to its action as a protective colloid. It is therefore useful for preparations containing tinctures of resinous materials that precipitate on addition to water. It is essential to ensure that any precipitated resin is well coated by the protective colloid before any electrolyte (which should be well diluted) is added. Acacia is not very effective for dense powders, and for these it is often combined with other thickeners such as tragacanth, starch and sucrose in compound tragacanth powder.

Unfortunately, acacia mucilage becomes acidic on storage as a result of enzyme activity, and it also contains an oxidase enzyme which may cause deterioration of active agents that are susceptible to oxidation. This enzyme can, however, be inactivated by heat.

Because of the stickiness of acacia it is rarely used in preparations for external use.

Tragacanth This product will form viscous aqueous solutions. Its thixotropic and pseudoplastic properties make it a better thickening agent than acacia and it can be used both for internal and external products. Like acacia it is mainly, though not exclusively, used for the extemporaneous preparation of suspensions with a short shelf-life.

Tragacanth is stable over a pH range of 4–7.5 but takes several days to hydrate fully after dispersion in water. The maximum viscosity of its dispersions is not, therefore, achieved until after this time, and can also be affected by heating. There are several grades of this material and only the best quality is suitable for use as a pharmaceutical suspending agent.

Alginates Alginic acid, a polymer of D-mannuronic acid, is prepared from kelp, and its salts have suspending properties similar to those of tragacanth. Alginate mucilages must not be heated above 60°C as depolymerization occurs, with a consequent loss in viscosity. They are most viscous immediately after preparation, after which there is a fall to a fairly constant value after about 24 hours. Alginates exhibit a maximum viscosity over a pH range of 5–9, and at low pH the acid is precipitated. Sodium alginate (Manucol) is the most widely used material in this class but it is, of course, anionic and will be incompatible with cationic materials and with heavy metals. The addition of calcium chloride to a sodium alginate dispersion will produce calcium alginate, which has a much higher viscosity. Several different viscosity grades are commercially available.

Starch Starch is rarely used on its own as a suspending agent but is one of the constituents of compound tragacanth powder, and it can also be used with carmellose sodium. Sodium starch glycollate (Explotab, Primojel), a derivative of potato starch, has also been evaluated for its use in the extemporaneous preparation of suspensions.

Xanthan gum (Keltrol) This is an anionic heteropolysaccharide produced by the action of *Xanthomonas campestris* on corn sugars. It is very soluble in cold water and is one of the most widely used thickening agents for the extemporaneous preparation of suspensions for oral use. It is used in concentrations up to about 2% and is stable over a wide pH range.

Water-soluble celluloses

Several cellulose derivatives are available that will disperse in water to produce viscous colloidal solutions suitable for use as suspending agents.

Methylcellulose (Celacol, Methocel) This is a semi-synthetic polysaccharide of the general formula:

$$[C_6H_7O_2(OH_2)OCH_3]_n$$

and is produced by the methylation of cellulose. Several grades are available, depending on their degree of methylation and on the chain length. The longer the chain, the more viscous is its solution. For example, a 2% solution of methylcellulose 20 exhibits an apparent viscosity of 20 millipascal seconds (mPa s) and methylcellulose 4500 has value of 4500 mPa s at 2% concentration. Because these products are more soluble in cold water than in hot, they are often dispersed in warm water and then, on cooling with constant stirring, a clear or opalescent viscous solution is produced. Methylcelluloses are non-ionic and therefore stable over a pH range of 3–11, and are compatible with many ionic additives. When these dispersions are heated, the methylcellulose molecules become progressively dehydrated and eventually gel at about 50°C; on cooling the original form is regained.

Hydroxyethylcellulose (Natrosol) This compound has hydroxyethyl instead of methyl groups attached to the cellulose chain and is also available in different viscosity grades. It has the advantage of being soluble in both hot and cold water and will not gel on heating. Otherwise it exhibits the same properties as methylcellulose.

Carmellose sodium (sodium carboxymethylcellulose) This material can be represented by:

$$[C_6H_{10-x}O_5(CH_2COONa)_x]_n$$

where x represents the degree of substitution, usually about 0.7, which in turn affects its solubility. The viscosity of its solution depends on the value of n, which represents the degree of polymerization. The numerical suffix gives an indication of the viscosity of a 2% solution. For example sodium carboxymethylcellulose 50 at a concentration of 2% will have a viscosity of 50 mPa s. This material produces clear solutions in both hot and cold water, which are stable over a pH range of about 5–10. Being anionic, this material is incompatible with polyvalent cations and the acid will be precipitated at low pHs. Heat sterilization of either the powder or its mucilage will reduce the viscosity, and this must be taken into account during formulation. It is widely used at concentrations of up to 1% in products for oral, parenteral or external use.

Microcrystalline cellulose This material consists of crystals of colloidal dimensions which disperse readily in water (but are not soluble) to produce thixotropic gels. It is a widely used suspending agent and the rheological properties of its dispersions can often be improved by the incorporation of additional hydrocolloid, in particular carboxymethylcellulose, methylcellulose and hydroxypropylmethylcellulose. These will aid dispersion and also stabilize the product against the flocculating effects of added electrolyte.

Hydrated silicates

There are three important materials within this classification, namely bentonite, magnesium aluminium silicate and hectorite, and they belong to a group called the montmorillonite clays. They hydrate readily, absorbing up to 12 times their weight of water, particularly at elevated temperatures. The gels formed are thixotropic and therefore have useful suspending properties. As with most naturally occurring materials they may be contaminated with spores, and this must be borne in mind when considering a sterilization process and choosing a preservative system.

Bentonite This has the general formula:

$$Al_2O_3.4SiO_2.H_2O$$

It is used at concentrations of up to 2 or 3% in preparations for external use, such as calamine lotion. As this product may contain pathogenic spores it should be sterilized before use.

Magnesium aluminium silicate (Veegum) Also known as attapulgite, this is available as insoluble flakes that disperse and swell readily in water by absorbing the aqueous phase into its crystal lattice. Several grades are available, differing in their particle size, their acid demand and the viscosity of their

dispersions. They can be used both internally and externally at concentrations of up to about 5%, and are stable over a pH range of 3.5–11. Veegum/water dispersions will exhibit thixotropy and plasticity with a high yield value, but the presence of salts can alter these rheological properties because of the flocculating effect of their positively charged counter-ions. Some grades, however, have a higher resistance to flocculation than others.

This material is often combined with organic thickening agents such as sodium carboxymethylcellulose or xanthan gum to improve yield values and degree of thixotropy, and to control flocculation (Ciullo 1981).

Hectorite This material is similar to bentonite and can be used at concentrations of 1–2% for external use. It is also possible to obtain synthetic hectorites (Laponite) that do not exhibit the batch variability or level of microbial contamination associated with natural products, and which can also be used internally.

As with other clays it is often advantageous to include an organic gum to modify its rheological properties.

Carbomers (carboxypolymethylene)

This material is a totally synthetic copolymer of acrylic acid and allyl sucrose. It is used at concentrations of up to 0.5%, mainly for external application, although some grades can be taken internally. When dispersed in water it forms acidic, low-viscosity solutions which, when adjusted to a pH of between 6 and 11, become highly viscous.

Colloidal silicon dioxide (Aerosil)

When dispersed in water this finely divided product will aggregate, forming a three-dimensional network. It can be used at concentrations of up to 4% for external use, but has also been used for thickening non-aqueous suspensions.

TYPES OF EMULSION

Pharmaceutical emulsions usually consist of a mixture of an aqueous phase with various oils and/or waxes. If the oil droplets are dispersed throughout the aqueous phase the emulsion is termed oil-in-water (o/w). A system in which the water is dispersed throughout the oil is a water-in-oil (w/o) emulsion. It is also possible to form multiple emulsions. For example, many small water droplets can be enclosed within larger oil droplets, which are themselves then dispersed in water. This gives a water-in-oil-in-water (w/o/w) emulsion. The alternative o/w/o emulsion is also possible.

If the dispersed globules are of colloidal dimensions (1 nm to 1 μm diameter) the preparation, which is quite often transparent or translucent, is called a microemulsion. This type has similar properties to a micellar system and will therefore exhibit the properties of hydrophobic colloids. As the size of the dispersed droplets increases more of the characteristics of coarse dispersions will be exhibited (see Chapter 6).

Tests for identification of emulsion type

Several simple methods are available for distinguishing between o/w and w/o emulsions (Table 23.1). The most common of these involve:

- miscibility tests with oil or water. The emulsion will only be miscible with liquids that are miscible with its continuous phase;
- conductivity measurements. Systems with aqueous continuous phases will readily conduct electricity, whereas systems with oily continuous phases will not;
- staining tests. Water-soluble and oil-soluble dyes are used, one of which will dissolve in, and colour the continuous phase.

FORMULATION OF EMULSIONS

Because of the very wide range of emulsifying agents available, considerable experience is required to choose the best emulgent system for a particular product. The final choice will depend to a large extent on the properties and use of the final product and the other materials required to be present.

Choice of emulsion type

The decision as to whether an o/w or a w/o emulsion is to be formulated will eliminate many unsuitable emulsifying systems.

Fats or oils for oral administration, either as medicaments in their own right or as vehicles for oil-soluble drugs, are invariably formulated as oil-in-water emulsions. In this form they are pleasant to take, and the inclusion of a suitable flavour in the aqueous phase will mask any unpleasant taste.

Table 23.1 Tests for identification of emulsion type

Oil-in-water emulsions	Water-in-oil emulsions
Miscibility tests	
Are miscible with water but immiscible with oil	Are miscible with oil but not with water
Staining tests by incorporation of an oil-soluble dye	
Macroscopic examination	
Paler colour than a w/o emulsion	More intense colouration than with an o/w emulsion
Microscopic examination	
Coloured globules on a colourless background	Colourless globules against a coloured background
Conductivity tests	
Water, being the continuous phase, will conduct electricity throughout the system. Two electrodes, when placed in such a preparation with a battery and suitable light source connected in series, will cause the lamp to glow	A preparation in which oil is the continuous phase will not conduct electricity. The lamp will not glow, or will only flicker spasmodically

Emulsions for intravenous administration must also be of the o/w type, although intramuscular injections can also be formulated as w/o products if a water-soluble drug is required for depot therapy.

Emulsions are most widely used for external application. Semisolid emulsions are termed creams and more fluid preparations are called either lotions or, if intended for massage into the skin, liniments. Both o/w and w/o types are available. The former is used for the topical application of water-soluble drugs, mainly for local effect. They do not have the greasy texture associated with oily bases and are therefore pleasant to use and easily washed from skin surfaces.

Water-in-oil emulsions will have an occlusive effect by hydration of the upper layers of the stratum corneum and the inhibition of evaporation of eccrine secretions. This, in turn, may influence the absorption rates of drugs from these preparations.

This type of emulsion is also useful for cleansing the skin of oil-soluble dirt, although its greasy texture is not always cosmetically acceptable. Oil-in-water emulsions are less efficient as cleansers but are usually more acceptable to the consumer, particularly for use on the hands. Similarly, moisturising creams, designed to prevent moisture loss from the skin and thus inhibit drying of the stratum corneum, are more efficient if formulated as w/o emulsions, which produce a coherent, water-repellent film.

Choice of oil phase

In many instances the oil phase of an emulsion is the active agent, and therefore its concentration in the product is predetermined. Liquid paraffin, castor oil, cod liver oil and arachis oil are all examples of medicaments which are formulated as emulsions for oral administration. Cottonseed oil, soya bean oil and safflower oil are used for their high calorific value in emulsions for intravenous feeding, and examples of externally applied oils that are formulated as emulsions include turpentine oil and benzyl benzoate.

Many emulsions for external use contain oils that are present as carriers for the active agent. It must be realized that the type of oil used may also have an effect both on the viscosity of the product and on the transport of the drug into the skin (see Chapter 33). One of the most widely used oils for this type of preparation is liquid paraffin. This is one of a series of hydrocarbons, which also includes hard paraffin, soft paraffin and light liquid paraffin. They can be used individually or in combination with each other to control emulsion consistency. This will ensure that the product can be spread easily but will be sufficiently viscous to form a coherent film over the skin. The film-forming capabilities of the emulsion can be further modified by the inclusion of various waxes, such as beeswax, carnauba wax or higher fatty alcohols. Continuous films can therefore be formed that are sufficiently tough and flexible to prevent contact between the skin and aqueous-based irritants. These preparations are called barrier creams, and many are of the w/o variety. The inclusion of silicone oils, such as dimethicone at 10–20%, which have exceptional water-repellent properties, may also permit the formulation of o/w products that are equally effective.

A variety of fixed oils of vegetable origin are also available, the most widely used being arachis, sesame, cottonseed and maize. Those expressed from seeds or fruits are often protein rich and contain useful vitamins and minerals. They are often, therefore, formulated for oral use as emulsions. Because of their lack of toxicity they can be used both internally and externally as vehicles for other materials.

Emulsion consistency

The texture or feel of a product intended for external use must also be considered. A w/o preparation will have a greasy texture and often exhibits a higher apparent viscosity than o/w emulsions. This fact is often used to convey a feeling of richness to many cosmetic formulations. Oil-in-water emulsions will, however, feel less greasy or sticky on application to the skin, will be absorbed more readily because of their lower oil content, and can be more easily washed from the skin surface.

Ideally emulsions should exhibit the rheological properties of plasticity/pseudoplasticity and thixotropy (see Chapter 4). A high apparent viscosity at the very low rates of shear caused by movement of dispersed phase globules is necessary in order to retard this movement and maintain a physically stable emulsion. It is important, however, that these products should flow freely when shaken, poured from the container or injected through a hypodermic needle. Therefore, at these high rates of shear, a lower apparent viscosity is required. This change in apparent viscosity must be reversible after a suitable time delay so as to retard creaming and coalescence.

For an externally applied product a wide range of emulsion consistencies can be tolerated. Low-viscosity lotions and liniments can be formulated that are dispensed from a flexible plastic container via a nozzle on to the skin. Only light shearing is then required to spread this type of product over the skin surface. This is particularly advantageous for painful or inflamed skin conditions.

The main disadvantage with low-viscosity emulsions is their tendency to cream easily, especially if formulated with a low oil concentration. It is rarely possible to formulate low-viscosity w/o products because of the consistency of the oil phase.

Emulsions of high apparent viscosity for external use are termed creams and are of a semisolid consistency. They are usually packed into collapsible plastic or aluminium tubes, although large volumes or very high-viscosity products are often packed into glass or plastic jars.

It is important not to ignore the patient/consumer acceptability of topically applied preparations, particularly in a competitive market.

There are several methods by which the rheological properties of an emulsion can be controlled.

Volume concentration of the dispersed phase

As discussed in Chapters 4 and 6, Einstein developed an equation relating the viscosity of a suspension to the volume fraction of the particles in that suspension. A qualitative application of this equation to the behaviour of emulsions shows that the viscosity of the product as a whole would be higher than the viscosity of the continuous phase on its own. So, as the concentration of dispersed phase increases, so does the apparent viscosity of the product.

Care must be taken to ensure that the dispersed phase concentration does not increase above about 60% of the total, as phase inversion may occur.

Particle size of the dispersed phase

It is possible, under certain conditions, to increase the apparent viscosity of an emulsion by a reduction in mean globule diameter. This can be achieved by homogenization. There are two postulated mechanisms for this occurrence:

1. A smaller mean globule size can cause increased flocculation. In a flocculated system a significant part of the continuous phase is trapped within aggregates of droplets, thus effectively increasing the apparent disperse phase concentration. Emulsions consisting of polydispersed droplets will tend to exhibit a lower viscosity than a monodispersed system, due to differences in electrical double-layer size and thus in the energy of interaction curves. These variations in interaction between globules during shear may be reflected in their flow behaviour,
2. If a hydrophilic colloid is used to stabilize the emulsion it will form a multimolecular film round the dispersed globules. A reduction in mean globule size will increase the total surface area, and therefore more colloid will be adsorbed on to the droplet surface. This will effectively increase the volume concentration of the dispersed phase.

The particle size of the dispersed phase is therefore controlled mainly by the method and conditions of manufacture of the emulsion, and by the type of emulgent used and its concentration.

Viscosity of the continuous phase

It has been well documented that a direct relationship exists between the viscosity of an emulsion and the viscosity of its continuous phase. Syrup and glycerol, which are used in oral emulsions as sweetening agents, will increase the viscosity of the continuous phase. Their main disadvantage is in increasing the density difference between the two phases, and thus possibly accelerating creaming.

Hydrocolloids, when used as emulsifying agents in o/w emulsions, will stabilize them not only by the formation of multimolecular layers around the dispersed globules, but also by increasing the continuous phase viscosity. They do not have the disadvantage of altering the density of this phase. If oil is the continuous phase, then the inclusion of soft or hard paraffin or certain waxes will increase its viscosity.

Viscosity of the dispersed phase

For most practical applications it is doubtful whether this factor would have any significant effect on total emulsion viscosity. It is possible, however, that a less viscous dispersed phase would, during shear, be deformed to a greater extent than a more viscous phase, and thus the total interfacial area would increase slightly. This may affect double-layer interactions and hence the viscosity of the emulsion.

Nature and concentration of the emulsifying system

It has already been shown that hydrophilic colloids, as well as forming multimolecular films at the oil/water interface, will also increase the viscosity of the continuous phase of an o/w emulsion. Obviously, as the concentration of this type of emulgent increases so will the viscosity of the product.

Surface-active agents forming condensed monomolecular films will, by the nature of their chemical structure, influence the degree of flocculation in a similar way, by forming linkages between adjacent globules and creating a gel-like structure. A flocculated system will exhibit a greater apparent viscosity than its deflocculated counterpart and will depend on surfactant concentration.

Choice of emulsifying agent

Toxicity and irritancy considerations

The choice of emulgent to be used will depend not only on its emulsifying ability, but also on its route of administration and, consequently, on its toxicity. Although there is no approved list of emulsifying agents for use in pharmaceutical products there is an approved list of emulsifiers as food additives for use in the European Union. It can be assumed that emulsifiers contained in this list would be suitable for internally used pharmaceutical emulsions. The regulations mainly include naturally occurring materials and their semisynthetic derivatives, such as the polysaccharides, as well as glycerol esters, cellulose ethers, sorbitan esters and polysorbates.

It will be noted that most of these are non-ionic, having a tendency to be less irritant and less toxic than their anionic, and particularly their cationic counterparts. The concentrations of ionic emulsifying agents necessary for emulsification will be irritant to the gastrointestinal tract and have a laxative effect, and should not be used for oral emulsions. Cationic surfactants in general are toxic even at lower concentrations. The emulgent cetrimide is limited to externally used preparations, where its antiseptic properties are of use.

Some emulgents, such as the anionic alkali soaps, often have a high pH and are thus unsuitable for application to broken skin. Even on normal intact skin with a pH of 5.5, the application of such alkaline materials can cause irritation. Some emulsifiers, in particular, wool fat can cause sensitization reactions in susceptible people.

When choosing an emulgent for parenteral use it must be realized that only certain types of non-ionic material are suitable. These include lecithin, polysorbate 80, methylcellulose, gelatin and serum albumin.

Formulation by the HLB method

It has already been shown that physically stable emulsions are best achieved by the presence of a condensed layer of emulgent at the oil/water interface, and that the complex interfacial films formed by a blend of an oil-soluble emulsifying agent with a water-soluble one produces the most satisfactory emulsions.

A useful method has been devised for calculating the relative quantities of these emulgents necessary to produce the most physically stable emulsion for a particular oil/water combination. This is called the hydrophile–lipophile balance (HLB) method. Although originally applied to non-ionic surface-active agents, its use has been extended to ionic emulgents. Each surfactant is allocated an HLB number representing the relative proportions of the lipophilic and hydrophilic parts of the molecule. High numbers (up to a theoretical maximum of 20) therefore indicate a surfactant exhibiting mainly hydrophilic or polar properties, whereas low numbers represent lipophilic or non-polar characteristics. Table 23.2 gives HLB values for some commonly used emulsifying agents. The concept of HLB values is discussed more fully in Chapter 6.

Each type of oil used will require an emulgent of a particular HLB number in order to ensure a stable

Table 23.2 HLB values for some pharmaceutical surfactants	
Sorbitan trioleate (Span 85)	1.8
Oleic acid	4.3
Sorbitan mono-oleate (Span 80)	4.3
Sorbitan monostearate (Span 60)	4.7
Sorbitan monolaurate (Span 20)	8.6
Polysorbate 60 (polyoxyethylene sorbitan monostearate)	14.9
Polysorbate 80 (polyoxyethylene sorbitan mono-oleate) (Tween 80)	15.0
Polysorbate 20 (polyoxyethylene sorbitan mono-laurate) (Tween 20)	16.7
Potassium oleate	20.0
Sodium dodecyl (lauryl) sulphate	40.0

product. For an o/w emulsion, for example, the more polar the oil phase the more polar must be the emulgent system.

Table 23.3 gives the required emulgent HLB value for particular oil phases for both types of emulsion. If a formulation contains a mixture of oils, fats or waxes the total HLB required can be calculated. The following example of an o/w emulsion will show this.

Liquid paraffin	35%
Wool fat	1%
Cetyl alcohol	1%
Emulsifier system	5%
Water	to 100%

The total percentage of oil phase is 37 and the proportion of each is:

Liquid paraffin	$35/37 \times 100$	=	94.6%
Wool fat	$1/37 \times 100$	=	2.7%
Cetyl alcohol	$1/37 \times 100$	=	2.7%

The total required HLB number is obtained as follows:

Table 23.3 Required HLB values for a range of oils and waxes		
	For a w/o emulsion	For an o/w emulsion
Beeswax	5	12
Cetyl alcohol	–	15
Liquid paraffin	4	12
Soft paraffin	4	12
Wool fat	8	10

Liquid paraffin (HLB 12)	$94.6/100 \times 12$ =	11.4
Wool fat (HLB 10)	$2.7/100 \times 10$ =	0.3
Cetyl alcohol (HLB 15)	$2.7/100 \times 15$ =	0.4
Total required HLB		**= 12.1**

From theoretical considerations, this particular formulation requires an emulgent blend of HLB 12.1 in order to produce the most stable emulsion. It must be realized, however, that the presence of other ingredients, particularly those that may partition into the oil phase, can also affect the required HLB value. It is therefore often necessary to prepare a series of emulsions using blends of a given pair of non-ionic emulsifying agents covering a wide range of HLB values. This is also important if the required HLB for an oil phase is not available. The HLB value of the emulgent blend giving the most stable emulsion is the required value for that oil phase.

Assuming that a blend of sorbitan mono-oleate (HLB 4.3) and polyoxyethylene sorbitan mono-oleate (HLB 15) is to be used as the emulsifying system, the proportions of each to be added to the emulsion to provide an HLB of 12.1 are calculated as follows.

Let A be the percentage concentration of the hydrophilic and B the percentage of the hydrophobic surfactants required to give a blend having an HLB value of x. Then:

$$A = \frac{100(x - \text{HLB of B})}{(\text{HLB of } A - \text{HLB of } B)} \quad \text{and} \quad B = 100 - A$$

In our example, therefore:

$$A = \frac{100(12.1 - 4.3)}{(15 - 4.3)} = 72.9$$

$B = 100 - 72.9 = 27.1$

Because the total percentage of emulgent blend in the formulation is 5, the percentage of each emulsifier will be:

Sorbitan mono-oleate	$5 \times 27.1/100 = 1.36$
Polyoxyethylene sorbitan mono-oleate	$5 - 1.36 = 3.64$

The series of trial emulsions can then be assessed for stability, based on the fact that the degree of creaming or separation is at a minimum at the optimal HLB value. Should several of the series show equally poor or equally good stability, resulting in an inability to choose a suitable HLB value, then the total emulgent concentration may be increased or reduced, respectively, and the manufacture of the series repeated.

Having determined the best HLB value for a given pair of emulgents, that value can now be used to

assess the suitability of other emulgent blends that may give a better emulsion than the one containing the emulgent used for the initial trials.

It must be remembered that, in choosing an emulsifier blend, the effect of chemical structure on the type of interfacial film must be taken into account. Condensed films are produced by emulgents having long, saturated hydrocarbon groups, thus providing maximum cohesion between adjacent molecules. In most cases it has been found that the most stable emulsions are formed when both emulsifying agents are of the same hydrocarbon chain length.

The use of phase inversion temperature

The use of the HLB system has several disadvantages, including the inability to take into account the effects of temperature, the presence of additives and the concentration of the emulsifier. It is possible to overcome some of these problems.

An o/w emulsion stabilized by non-ionic emulgents will, on heating, invert to form a w/o product. This is because, as the temperature increases, the HLB value of a non-ionic surfactant will decrease as it becomes more hydrophobic. At the temperature at which the emulgent has equal hydrophilic and hydrophobic tendencies (the phase inversion temperature) the emulsion will invert.

The stability of an emulsion has been related to the phase inversion temperature (PIT) of its emulsifying agent (see Chapter 6).

CLASSIFICATION OF EMULSIFYING AGENTS

The inclusion of an emulsifying agent or agents is necessary to facilitate actual emulsification during manufacture, and also to ensure emulsion stability during the shelf-life of the product.

The different methods by which emulsifying agents (also called emulsifiers or emulgents) exert their effects have been detailed in Chapter 6, but the one factor common to all of them is their ability to form an adsorbed film around the dispersed droplets between the two phases. There are many types of emulgent available, but for convenience they can be divided into two main classifications: synthetic or semisynthetic surface-active agents, and naturally occurring materials and their derivatives.

These divisions are quite arbitrary and some materials may justifiably be placed in more than one category.

Synthetic and semisynthetic surface-active agents

There are four main categories of these materials, depending on their ionization in aqueous solutions: anionic, cationic, non-ionic and amphoteric.

Anionic surfactants

In aqueous solutions these compounds dissociate to form negatively charged anions that are responsible for their emulsifying ability. They are widely used because of their cheapness, but because of their toxicity are only used for externally applied preparations.

Alkali metal and ammonium soaps Emulgents in this group consist mainly of the sodium, potassium or ammonium salts of long-chain fatty acids, such as:

sodium stearate $C_{17}H_{35}COO^- Na^+$

They produce stable o/w emulsions but may in some instances require the presence of an auxiliary non-ionic emulsifying agent in order to form a complex monomolecular film at the oil/water interface. Because in acidic conditions, these materials will precipitate out as the free fatty acids, they are most efficient in an alkaline medium.

This type of emulgent can also be formed in situ during the manufacture of the product by reacting an alkali such as potassium, sodium or ammonium hydroxide with a fatty acid. The latter may be a constituent of a vegetable oil. Oleic acid and ammonia, for example, are reacted together to form the soap responsible for stabilizing White Liniment.

These emulgents are incompatible with polyvalent cations, often causing phase reversal, and it is therefore essential that deionized water is used in their preparation.

Soaps of divalent and trivalent metals Although many different divalent and trivalent salts of fatty acids exist, and would produce satisfactory emulsions, only the calcium salts are commonly used. They are often formed in situ during preparation of the product by interacting the appropriate fatty acid with calcium hydroxide. For example, oleic acid is reacted with calcium hydroxide to produce calcium oleate, which is the emulsifying agent for both Zinc Cream BP and some formulations of oily calamine lotion.

These emulgents will only produce w/o emulsions.

Amine soaps A number of amines form salts with fatty acids. One of the most important of those used is based on triethanolamine $N(CH_2CH_2OH)_3$ and is widely used in both pharmaceutical and cosmetic products. For example, triethanolamine stearate

forms stable o/w emulsions and is usually made in situ by a reaction between triethanolamine and the appropriate fatty acid. Although these emulgents are usually pH neutral they are still restricted to externally used preparations. They are also incompatible with acids and high concentrations of electrolytes.

Sulphated and sulphonated compounds The alkyl sulphates have the general formula $ROSO_3^-M^+$, where R represents a hydrocarbon chain and M^+ is usually sodium or triethanolamine. An example is sodium lauryl sulphate, which is widely used to produce o/w emulsions. Because of its high water solubility and its inability to form condensed films at the oil/water interface, it is always used in conjunction with a non-ionic oil-soluble emulsifying agent in order to produce a complex condensed film. It is used with cetostearyl alcohol to produce Emulsifying Wax, which stabilizes such preparations as Aqueous Cream and Benzyl Benzoate Application.

Sulphonated compounds are much less widely used as emulgents. Materials of this class include sodium dioctylsulphosuccinate, and are more often used as wetting agents or for their detergency.

Cationic surfactants

In aqueous solutions these materials dissociate to form positively charged cations that provide the emulsifying properties. The most important group of cationic emulgents consists of the quaternary ammonium compounds. Although these materials are widely used for their disinfectant and preservative properties, they are also useful o/w emulsifiers. Like many anionic emulgents, if used on their own they will produce only poor emulsions, but if used with non-ionic oil-soluble auxiliary emulgents they will form stable preparations.

Because of the toxicity of cationic surfactants they tend to be used only for the formulation of antiseptic creams, where the cationic nature of the emulgent is also responsible for the product's antiseptic properties.

Cationic emulsifying agents are incompatible with anionic surface-active agents and polyvalent anions, and are unstable at high pH.

Cetrimide The most useful of these cationic emulgents is cetrimide (cetyl trimethylammonium bromide) $CH_3(CH_2)_{15}N^+(CH_3)_3Br^-$. Cetrimide is used at a concentration of 0.5% with 5% cetostearyl alcohol for the formulation of Cetrimide Cream BP.

Non-ionic surfactants

These products range from oil-soluble compounds stabilizing w/o emulsions to water-soluble materials giving o/w products. It is usual for a combination of a water-soluble with an oil-soluble emulgent to be used in order to obtain the complex interfacial film necessary for optimum emulsion stability. Non-ionic emulgents are particularly useful because of their low toxicity and irritancy; some can therefore be used for orally and parenterally administered preparations. They also have a greater degree of compatibility with other materials than do anionic or cationic emulgents, and are less sensitive to changes in pH or to the addition of electrolytes. They do, however, tend to be more expensive.

Being non-ionic, the dispersed globules may not possess a significant charge density. To reduce the tendency for coalescence to occur in an oil-in-water emulsion, it is necessary that the polar groups be well hydrated and/or sufficiently large to prevent close approach of the dispersed droplets in order to compensate for the lack of charge.

Most non-ionic surfactants are based on:

- a fatty acid or alcohol (usually with 12–18 carbon atoms), the hydrocarbon chain of which provides the hydrophobic moiety;
- an alcohol (-OH) and/or ethylene oxide grouping ($-OCH_2CH_2-$), which provide the hydrophilic part of the molecule.

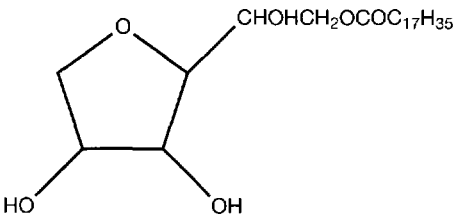
By varying the relative proportions of the hydrophilic and hydrophobic groupings many different products can be obtained.

If the hydrophobic part of the molecule predominates, then the surfactant will be oil-soluble. It will not concentrate at the oil/water interface but rather tend to migrate into the oil phase. Similarly, a water-soluble surfactant will migrate into the aqueous phase and away from the oil/water interface. The best type of non-ionic surfactant to use is one with an equal balance of hydrophobic and hydrophilic groupings. An alternative would be to use two emulgents, one hydrophilic and one hydrophobic. The cohesion between their hydrocarbon chains will then hold both types at the oil/water interface.

Glycol and glycerol esters Glyceryl monostearate (a polyhydric alcohol fatty acid ester) is a strongly hydrophobic material that produces weak w/o emulsions. The addition of small amounts of sodium, potassium or triethanolamine salts of suitable fatty acids will produce a 'self-emulsifying' glyceryl monostearate, which is a useful o/w emulsifier. Self-emulsifying monostearin is glyceryl monostearate to which anionic soaps (usually oleate or stearate) have been added. This combination is used to stabilize Hydrocortisone Lotion.

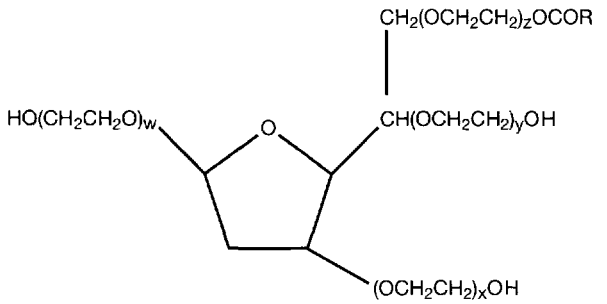
Other polyhydric alcohol fatty acid esters are also available either in the pure form or in the 'self-emulsifying' form containing small proportions of a primary emulsifier, and include glyceryl mono-oleate, diethylene glycol monostearate and propylene glycol mono-oleate.

Sorbitan esters These are produced by the esterification of one or more of the hydroxyl groups of sorbitan with either lauric, oleic, palmitic or stearic acids. The structure of sorbitan monostearate is shown below.



This range of surfactants exhibits lipophilic properties and tends to form w/o emulsions. They are, however, much more widely used with polysorbates to produce either o/w or w/o emulsions.

Polysorbates Polyethylene glycol derivatives of the sorbitan esters give us polysorbates. These have the general formula:



where R represents a fatty acid chain. Variations in the type of fatty acid used and in the number of oxyethylene groups in the polyethylene glycol chains produce a range of products of differing oil and water solubilities. Polyoxyethylene 20 sorbitan mono-oleate, for example, contains 20 oxyethylene groups in the molecule. This number must not be confused with the one given as part of the official name (Polysorbate 80) or in the trade name (Tween 80), which is included in order to identify the type of fatty acid in the molecule.

Polysorbates are generally used in conjunction with the corresponding sorbitan ester to form a complex condensed film at the oil/water interface (see Formulation by the HLB method, earlier).

Other non-ionic oil-soluble materials, such as glyceryl monostearate, cetyl or stearyl alcohol or propylene glycol monostearate, can be incorporated with polysorbates to produce 'self-emulsifying' preparations. For example, Polawax contains cetyl alcohol with a polyoxyethylene sorbitan ester.

Polysorbates are compatible with most anionic, cationic and non-ionic materials. They are pH neutral and are stable to the effects of heat, pH change and high concentrations of electrolyte. Their low toxicity renders them suitable for oral use and some are also used in parenteral preparations. They have the disadvantage, however, of an unpleasant taste, and care must be taken when selecting a suitable preservative as many are inactivated by complexation with polysorbates.

Fatty alcohol polyglycol ethers These are condensation products of polyethylene glycol and fatty alcohols, usually cetyl or cetostearyl:

$$ROH + (CH_2CH_2O)_n \rightarrow RO(CH_2CH_2O)_nH$$

where R is a fatty alcohol chain.

Perhaps the most widely used is macrogol cetostearyl ether (22) or cetomacrogol 1000, which is polyethylene glycol monocetyl ether. This is a very useful water-soluble o/w emulgent, but because of its high water solubility it is necessary to include an oil-soluble auxiliary emulsifier when formulating emulsions. Cetomacrogol Emulsifying Ointment includes cetomacrogol 1000 and cetostearyl alcohol and is used to stabilize cetomacrogol creams.

They can also be produced with shorter polyoxyethylene groups as lipophilic w/o emulsifiers. Combinations of lipophilic and hydrophilic ethers can be used together to produce stable emulsions.

These materials can be salted out by the addition of high concentrations of electrolyte, but are stable over a wide pH range.

Fatty acid polyglycol esters The stearate esters or polyoxyl stearates are the most widely used of this type of emulgent. Polyoxyethylene 40 stearate (in which 40 represents the number of oxyethylene units) is a water-soluble material often used with stearyl alcohol to give o/w emulsions.

Poloxalkols Poloxalkols are polyoxyethylene/polyoxypropylene copolymers with the general formula:

$$OH(C_2H_4O)_a(C_3H_6O)_b(C_2H_4O)_a$$

and comprise a very large group of compounds, some of which are used as emulsifying agents for intravenous fat emulsions.

Higher fatty alcohols The hexadecyl (cetyl) and octadecyl (stearyl) members of this series of saturated aliphatic monohydric alcohols are useful auxiliary emulsifying agents. Part of their stabilizing effect

comes from their ability to increase the viscosity of the preparation, thereby retarding creaming. Cetostearyl alcohol will also form complex interfacial films with hydrophilic surface-active agents such as sodium lauryl sulphate, cetrimide or cetomacrogol 1000, and so stabilize o/w emulsions.

Amphoteric surfactants

This type possesses both positively and negatively charged groups, depending on the pH of the system. They are cationic at low pH and anionic at high pH. Although they are not widely used as emulsifying agents, one example, lecithin, is used to stabilize intravenous fat emulsions.

Naturally occurring materials and their derivatives

Naturally occurring materials often suffer from two main disadvantages: they show considerable batch-to-batch variation in composition and hence in emulsifying properties, and many are susceptible to bacterial or mould growth. For these reasons they are not widely used in manufactured products requiring a long shelf-life, but rather for extemporaneously prepared emulsions designed for use within a few days of manufacture.

Polysaccharides

The most important emulsifying agent in this group is acacia. This stabilizes o/w emulsions by forming a strong multimolecular film round each oil globule, and so coalescence is retarded by the presence of a hydrophilic barrier between the oil and water phases.

Because of its low viscosity, creaming will occur readily, and therefore a suspending agent such as tragacanth or sodium alginate can also be included. Because of its sticky nature the use of acacia is limited to products for internal use.

Semisynthetic polysaccharides

In order to reduce the problems associated with batch-to-batch variation, semisynthetic derivatives are available as o/w emulgents or stabilizers.

Several grades of methylcellulose and carmellose sodium are available and exert their action in a similar way to that of acacia.

Methylcellulose 20, for example, is used at a concentration of 2% to stabilize Liquid Paraffin Oral Emulsion.

Sterol-containing substances

Beeswax, wool fat and wool alcohols are all used in the formulation of emulsions. Beeswax is used mainly in cosmetic creams of both o/w and w/o type, in conjunction with borax. Because of the systemic toxicity of boric acid and its salts, however, the use of beeswax/borax preparations is limited, although beeswax is used as a stabilizer for w/o creams.

Wool fat (anhydrous lanolin) consists chiefly of normal fatty alcohols with fatty acid esters of cholesterol and other sterols. It will form w/o emulsions of low dispersed phase concentration, and it can also be incorporated for its emollient properties. Some individuals exhibit sensitization to this material and, because of its characteristic odour and the need to incorporate antioxidants, it is not widely used. It is, however, to be found in low concentrations in many ointments, where its water-absorbing properties are of great value. It can be employed as an emulsion stabilizer with a primary emulsifying agent, for example with calcium oleate in oily calamine lotion, with beeswax in Proflavine Cream, and with cetostearyl alcohol in Zinc Cream and ichthammol cream.

Because wool fat has some ideal properties, attempts have been made to improve its other, less desirable, characteristics by physical and chemical modification. Processes including hydrogenation and fractionation have been carried out with some success. It has also been converted, by a reaction with ethylene oxide, to give a range of polyoxyethylene lanolin derivatives. These non-ionic products are mainly water soluble and are used as o/w emulgents possessing the properties of emollience.

The principal emulsifying agent in wool fat is wool alcohols, which consists mainly of cholesterol together with other alcohols. It is an effective w/o emulgent, being more powerful than wool fat, and is used in the formulation of Hydrous Ointment. It is also incorporated as Wool Alcohols Ointment into other ointment bases which, although not emulsions, will readily mix with aqueous skin secretions and easily wash off the skin. Wool alcohols does not have the same strong odour as wool fat but does require the presence of an antioxidant.

Finely divided solids

Certain finely divided solids can be adsorbed at the oil/water interface, forming a coherent film that physically prevents coalescence of the dispersed globules. If the particles are preferentially wetted by the aqueous phase then o/w products will result, whereas preferential wetting by the oil will produce w/o emulsions.

Montmorillonite clays (such as bentonite and aluminium magnesium silicate) and colloidal silicon dioxide are used mainly for external use. Aluminium and magnesium hydroxides are also used internally. For example, Liquid Paraffin and Magnesium Hydroxide Oral Emulsion BP is stabilized by the incorporation of the magnesium hydroxide.

Other formulation additives

Buffers

The inclusion of buffers (see Chapters 3, 21 and 35) may be necessary to maintain chemical stability, control tonicity or ensure physiological compatibility. It must be remembered, however, that the addition of electrolytes may have profound effects on the physical stability of suspensions and emulsions.

Density modifiers

From a qualitative examination of Stokes' law (Chapter 6) it can be seen that if the disperse and continuous phases both have the same densities then sedimentation or creaming will not occur. Minor modifications to the aqueous phase of a suspension or emulsion by incorporating sucrose, dextrose, glycerol or propylene glycol can be achieved, but because of the differing coefficients of expansion this can only be possible over a small temperature range.

Humectants

Glycerol, polyethylene glycol and propylene glycol are examples of suitable humectants that can be incorporated at concentrations of about 5% into aqueous suspensions for external application. They are used to prevent the product from drying out after application to the skin.

They can also be added to an emulsion formulation in order to reduce the evaporation of the water, either from the packaged product when the closure is removed or from the surface of the skin after application. High concentrations, if used topically, may actually remove moisture from the skin, thereby dehydrating it.

Antioxidants

Before including an antioxidant in emulsion formulations, it is essential to ensure that its use is not restricted in whichever country it is desired to sell the product. In Britain, butylated hydroxyanisole (BHA) is widely used for the protection of fixed oils and fats at concentrations of up to 0.02% and for some essential oils up to 0.1%. A similar antioxidant is butylated hydroxytoluene (BHT), which is recommended as an alternative to tocopherol at a concentration of 10 ppm to stabilize liquid paraffin. Other antioxidants widely used for emulsion formulation include the propyl, octyl and dodecyl esters of gallic acid, recommended for use at concentrations up to 0.001% for fixed oils and fats and up to 0.1% for essential oils.

The efficiency of an antioxidant in a product will depend on many factors, including its compatibility with other ingredients, its oil/water partition coefficient, the extent of its solubilization within micelles of the emulgent, and its sorption on to the container and its closure. It must be realized, therefore, that the choice of antioxidant and the concentration at which it is to be used can only be determined by testing its effectiveness in the final product and in the package in which the product is to be sold.

Flavours, colours and perfumes

The use of these ingredients is discussed in Chapter 21 and the information there will be directly applicable to suspension and emulsion formulation.

Adsorption of these materials on to the surfaces of the dispersed phase of a suspension may occur, and because of the high surface area of the dispersed powders in this type of formulation, their effective concentrations in solution may be significantly reduced. The finer the degree of subdivision of the disperse phase, the paler may appear the colour of the product for a given concentration of dye.

It must also be realized that the inclusion of these adjuvants may alter the physical characteristics of both suspensions and emulsions. Either the presence of electrolytes or their effect on pH can influence the degree of flocculation.

Sweetening agents

Suitable sweeteners are discussed in Chapter 21. High concentrations of sucrose, sorbitol or glycerol, which will exhibit Newtonian properties, may adversely affect the rheological properties of the suspension. Synthetic sweeteners may be salts and can affect the degree of flocculation.

Preservation of suspensions and emulsions

Preservation of suspensions

The section covering the preservation of emulsions is applicable also to suspension formulation. It is

essential that a suitable preservative be included, particularly if naturally occurring materials are to be used. This is to prevent the growth of micro-organisms that may be present in the raw material and/or introduced into the product during use. Some of the natural products, particularly if they are to be applied to broken skin, should be sterilized before use. Bentonite, for example, may contain *Clostridium tetani* but can be sterilized by heating the dry powder at 160°C for 1 hour or by autoclaving aqueous dispersions.

As with emulsion formulation, care must be taken to ascertain the extent of inactivation, if any, of the preservative system caused by interaction with other excipients. Solubilization by wetting agents, interaction with polymers or adsorption on to suspended solids, particularly kaolin or magnesium trisilicate, may reduce the availability of preservatives.

Preservation of emulsions

Problems associated with the growth of micro-organisms in pharmaceutical products are discussed in Chapter 43. Those microbiological factors of specific importance to the stability of emulsions are discussed later in this chapter. The necessity of including a preservative in an emulsion formulation is discussed below.

Unfortunately there is no theoretical way of choosing a suitable preservative system, the only reliable methods being based on the results of suitable challenge tests. These methods of testing preservative activity are given in official compendia, but essentially involve the addition to the test products of a mixture of Gram-positive and Gram-negative bacteria, yeasts and moulds, and comparing their survival with a control sample containing no preservative.

The desirable features of a preservative suitable for use in an emulsion include:

- a wide spectrum of activity against all bacteria, yeasts and moulds;
- bactericidal rather than bacteristatic activity. A preservative having a minimal bacteristatic activity may lose it if any physical or chemical changes occur in the system;
- freedom from toxic, irritant or sensitizing activity;
- high water solubility. Because the growth of microorganisms occurs in the aqueous phase, it is important that the preservative has a low oil/water partition coefficient. The more polar the oil phase, the more difficult it is to preserve the product adequately, owing to the solubility of the preservative in both phases. If the preservative is more soluble in oil than in water, then increasing the proportion of oil will decrease the aqueous phase concentration. Allowance must be made for this when choosing the phase–volume ratios;
- compatibility with the other ingredients and with the container. Certain preservatives are incompatible with particular groups of emulsifying agent. Phenols and the esters of *p*-hydroxybenzoic acid, for example, will complex with some non-ionic emulgents, owing to either a reaction with oxyethylene groups or solubilization within micelles of excess surfactant. In many cases it is possible, by chemical assay, to detect the correct concentration of preservative in the product even though some of it may not be available for antimicrobial activity. If some of the added preservative has been inactivated it may be possible to overcome this problem by increasing the amount of preservative in the product to give a satisfactory concentration of free preservative in the aqueous phase. It is important to ensure that, during manufacture, the preservative is added after the emulgent has concentrated at the oil/water interface;
- stability and effectiveness over a wide range of pH and temperatures;
- freedom from colour and odour;
- retention of activity in the presence of large numbers of microorganisms. Uptake of preservative by bacterial cells may deplete the concentration of preservative in solution, thereby rendering it insufficient to maintain adequate bactericidal activity.

Because of the complex systems involved and the many factors to be taken into consideration, it is necessary to test the efficiency of a new preservative in the finished product and container by suitable challenge testing procedures.

The most widely used preservatives in emulsions include benzoic and sorbic acid and their salts, *p*-hydroxybenzoic acid esters, chlorocresol, phenoxyethanol, bronopol, quaternary ammonium compounds and, to a lesser extent, organic mercurials. Because of the irritancy and toxicity of certain preservatives, the initial choice will depend on the route of administration of the product. Further details of the use of preservatives in emulsions can be found in Chapter 42.

It must be realized that no single preservative exhibits all of the desirable properties outlined earlier. In many cases a combination is required, the most widely used being a mixture of methyl and propyl *p*-hydroxybenzoates at a ratio usually of 10:1.

PHYSICAL STABILITY OF SUSPENSIONS

The physical stability of a suspension is normally assessed by the measurement of its rate of sedimentation, the final volume or height of the sediment, and the ease of redispersion of the product.

The first two parameters can be assessed easily by a measurement of the total initial volume or height of the suspension (V_0) and the volume or height of the sediment (V_t), as shown in Figure 23.1. By plotting the value of V_t/V_0 against time for a series of trial formulations (all initial values will equal unity), it can be seen, by an assessment of the slope of each line, which suspension shows the slowest rate of sedimentation. When the value of V_t/V_0 becomes constant this indicates that sedimentation has ceased.

Alternatively, the term flocculation value can be used, which is a ratio of the final volume or height of the sediment and the volume or height of the fully sedimented cake of the same system which has been deflocculated.

Attempts have also been made to equate the zeta potential of the suspended particles with the physical stability – particularly the degree of flocculation – of the system using electrophoresis.

The ease of redispersion of the product can be assessed qualitatively by simply agitating the product in its container. The use of a mechanical shaker will eliminate variations in shaking ability.

PHYSICAL STABILITY OF EMULSIONS

A stable emulsion is one in which the dispersed globules retain their initial character and remain uniformly distributed throughout the continuous phase. Various types of deviation from this ideal behaviour can occur. Explanations for emulsion stability have been given in Chapter 6. This section will concentrate on methods of improving emulsion stability in practice.

Creaming and its avoidance

This is the separation of an emulsion into two regions, one of which is richer in the disperse phase than the other. A simple example is the creaming of milk, when fat globules slowly rise to the top of the product. This is not a serious instability problem as a uniform dispersion can be reobtained simply by shaking the emulsion. It is, however, undesirable because of the increased likelihood of coalescence of the droplets, owing to their close proximity to each other. A creamed emulsion is also inelegant and, if the emulsion is not shaken adequately, there is a risk of the patient obtaining an incorrect dosage.

Consideration of the qualitative application of Stokes' law will show that the rate of creaming can be reduced by the following methods.

Production of an emulsion of small droplet size

This factor usually depends on the method of manufacture. An efficient emulsifying agent will not only stabilize the emulsion but also facilitate the actual emulsification process to give a product of fine globule size.

Increase in the viscosity of the continuous phase

Many auxiliary emulsifying agents, in particular the hydrophilic colloids, are viscosity enhancers and this property is part of their emulsifying capability. For example, the inclusion of methylcellulose will reduce the mobility of the dispersed droplets in an o/w emulsion. The addition of soft paraffin will have the same effect on water droplets in a w/o emulsion.

Storage of the product at a low temperature (but above freezing point) will increase the viscosity of the continuous phase and also reduce the kinetic energy of the system. This will decrease the rate of migration of the globules of the disperse phase. It is unwise, however, to rely solely on this method of controlling creaming, as storage conditions after the product is sold are outside the control of the manufacturer.

Reduction in the density difference between the two phases

Creaming could be prevented altogether if the densities of the two phases were identical. In practice this method is never used, as it could only be achieved over a very narrow temperature range owing to differences in the coefficients of expansion between different ingredients.

Control of disperse phase concentration

It is not easy to stabilize an emulsion containing less than 20% disperse phase, as creaming will readily occur. A higher disperse phase concentration would result in a hindrance of movement of the droplets and hence in a reduction in rate of creaming.

Although it is theoretically possible to include as much as 74% of an internal phase, it is usually found that at about 60% concentration phase inversion occurs.

Finally, it must be realized that some of the factors above are interrelated. For example, homogenization of the emulsion would decrease globule size and, by thus increasing their number, increase the viscosity of the product.

Flocculation prevention

Flocculation involves the aggregation of the dispersed globules into loose clusters within the emulsion. The individual droplets retain their identities but each cluster behaves physically as a single unit. This, as we have already seen, would increase the rate of creaming. As flocculation must precede coalescence, any factor preventing or retarding flocculation would therefore maintain the stability of the emulsion.

Flocculation in the secondary minimum (see Fig. 6.3) occurs readily and cannot be avoided. Redispersion can easily be achieved by shaking. Primary minimum flocculation, however, is more serious and redispersion is not so easy.

The presence of a high charge density on the dispersed droplets will ensure the presence of a high energy barrier, and thus reduce the incidence of flocculation in the primary minimum. Care must be taken to ensure that the effects of any ions in the product are taken into consideration very early in the formulation process. This is particularly important when formulating emulsions for parenteral nutrition which contain high levels of electrolytes (Washington 1990).

Coalescence (breaking, cracking)

The coalescence of oil globules in an o/w emulsion is resisted by the presence of a mechanically strong adsorbed layer of emulsifier around each globule. This is achieved by the presence of either a condensed mixed monolayer of lipophilic and hydrophilic emulgents, or a multimolecular film of a hydrophilic material. Hydration of either of these types of film will hinder the drainage of water from between adjacent globules which is necessary prior to coalescence. As two globules, approach each other their close proximity causes their adjacent surfaces to flatten. As a change from a sphere to any other shape results in an increase in surface area and hence in total surface free energy, this globule distortion will be resisted and drainage of the film of continuous phase from between the two globules will be delayed.

The presence of long, cohesive hydrocarbon chains projecting into the oil phase will prevent coalescence in a w/o emulsion.

CHEMICAL INSTABILITY OF EMULSIONS

Although it is not possible to list every incompatibility, the following general points will illustrate the more common chemical problems that can cause the coalescence of an emulsion.

It is necessary to ensure that any emulgent system used is not only physically but also chemically compatible with the active agent and with the other emulsion ingredients. Ionic emulsifying agents, for example, are usually incompatible with materials of opposite charge. Anionic and cationic emulgents are thus mutually incompatible.

It has already been demonstrated that the presence of electrolyte can influence the stability of an emulsion either by:

- reducing the energy of interaction between adjacent globules, or
- a salting-out effect, by which high concentrations of electrolytes can strip emulsifying agents of their hydrated layers and so cause their precipitation.

In some cases phase inversion may occur rather than demulsification. If, for example, a sodium soap is used to stabilize an o/w emulsion, then the addition of a divalent electrolyte such as calcium chloride may form the calcium soap, which will stabilize a w/o emulsion.

Emulgents may also be precipitated by the addition of materials in which they are insoluble. It may be possible to precipitate hydrophilic colloids by the addition of alcohol. For this reason care must therefore be taken if tinctures are to be included in emulsion formulations.

Changes in pH may also lead to the breaking of emulsions. Sodium soaps may react with acids to produce the free fatty acid and the sodium salt of the acid. Soap-stabilized emulsions are therefore usually formulated at an alkaline pH.

Oxidation

Many of the oils and fats used in emulsion formulation are of animal or vegetable origin and can be susceptible to oxidation by atmospheric oxygen or by the action of microorganisms. The resulting rancidity is manifested by the formation of

degradation products of unpleasant odour and taste. These problems can also occur with certain emulsifying agents, such as wool fat or wool alcohols. Oxidation of microbiological origin is controlled by the use of antimicrobial preservatives, and atmospheric oxidation by the use of reducing agents or, more usually, antioxidants. Some examples are mentioned earlier in this chapter.

Microbiological contamination

The contamination of emulsions by microorganisms can adversely affect the physicochemical properties of the product, causing such problems as gas production, colour and odour changes, hydrolysis of fats and oils, pH changes in the aqueous phase, and breaking of the emulsion. Even without visible signs of contamination an emulsion can contain many bacteria and, if these include pathogens, may constitute a serious health hazard. Most fungi and many bacteria will multiply readily in the aqueous phase of an emulsion at room temperature, and many moulds will also tolerate a wide pH range. Some of the hydrophilic colloids, which are widely used as emulsifying agents, may provide a suitable nutritive medium for use by bacteria and moulds. Species of the genus *Pseudomonas* can utilize polysorbates, aliphatic hydrocarbons and compounds. Some fixed oils, including arachis oil, can be used by some *Aspergillus* and *Rhizopus* species, and liquid paraffin by some species of *Penicillium*.

A few emulgents, particularly those from natural sources, may introduce heavy contamination into products in which they are used. Because bacteria can reproduce in resin beds, deionized water may be unsatisfactory and even distilled water, if incorrectly stored after collection, can be another source of contamination. Oil-in-water emulsions tend to be more susceptible to microbial spoilage than water-in-oil products as, in the latter case the continuous oil phase acts as a barrier to the spread of microorganisms throughout the product, and the less water there is present the less growth there is likely to be.

It is therefore, necessary to include an antimicrobial agent to prevent the growth of any microorganisms that might contaminate the product. Suitable candidates are discussed in Chapter 42.

Adverse storage conditions

Adverse storage conditions may also cause emulsion instability. It has already been explained that an increase in temperature will cause an increase in the rate of creaming, owing to a fall in apparent viscosity of the continuous phase. The temperature increase will also cause an increased kinetic motion, both of the dispersed droplets and of the emulsifying agent at the oil/water interface. This effect on the disperse phase will enable the energy barrier to be easily surmounted and thus the number of collisions between globules will increase. Increased motion of the emulgent will result in a more expanded monolayer, and so coalescence is more likely. Certain macromolecular emulsifying agents may also be coagulated by an increase in temperature.

At the other extreme, freezing of the aqueous phase will produce ice crystals that may exert unusual pressures on the dispersed globules and their adsorbed layer of emulgent. In addition, dissolved electrolyte may concentrate in the unfrozen water, thus affecting the charge density on the globules. Certain emulgents may also precipitate at low temperatures.

The growth of microorganisms within the emulsion can cause deterioration and it is therefore essential that these products are protected as far as possible from the ingress of microorganisms during manufacture, storage and use, and that they contain adequate preservatives.

STABILITY TESTING OF EMULSIONS

Methods of assessing stability

Macroscopic examination

The physical stability of an emulsion can be assessed by an examination of the degree of creaming or coalescence occurring over a period of time. This is carried out by calculating the ratio of the volume of the creamed or separated part of the emulsion and the total volume. These values can be compared for different products.

Globule size analysis

If the mean globule size increases with time (coupled with a decrease in globule numbers), it can be assumed that coalescence is the cause. It is therefore possible to compare the rates of coalescence for a variety of emulsion formulations by this method. Microscopic examination or electronic particle counting devices, such as the Coulter counter, or laser diffraction sizing are most widely used.

Viscosity changes

It has already been shown that many factors influence the viscosity of emulsions. Any variation in globule size or number, or in the orientation or migration of emulsifier over a period of time, may be detected by a change in apparent viscosity. Suitable methods and equipment are detailed in Chapter 4.

In order to compare the relative stabilities of a range of similar products it is often necessary to speed up the processes of creaming and coalescence. This can be achieved in one of the following ways.

Accelerated stability tests

To assess the physical stability of suspensions and emulsions macroscopic examination and measurement of apparent viscosity are of value. In addition, for emulsions, microscopic evaluation of globule size distribution and numbers will provide further evidence of changes in physical stability.

Storage at adverse temperatures

An assessment of these parameters at elevated temperatures for emulsions and coarse suspensions would give a speedier indication of a rank order of degree of instability, but it is essential to correlate these results with those taken from suspensions stored at ambient temperatures.

Temperature cycling By exaggerating the temperature fluctuations to which any product is subjected under normal storage conditions, it may be possible to compare the physical stabilities of a series of suspensions or emulsions. Temperature cycles consisting of storage for several hours at about 40°C, followed by refrigeration or freezing until instability becomes evident, have been used successfully. The continual formation and melting of small ice crystals will disrupt the adsorbed layer of emulgent at the oil/water interface, and any weakness in the structure of the film will quickly become apparent. Similarly, normal temperature fluctuations can be used, but at increased frequencies of only a few minutes at each extreme. This method of accelerated stability testing is particularly useful for the assessment of crystal growth in suspensions. Measurement of particle size is usually carried out microscopically, by laser diffraction or by use of a Coulter counter. It is of course important to ensure that the suspension is deflocculated to ensure that each individual particle is measured, rather than each floccule.

Centrifugation

A qualitative examination of Stokes' law (Chapters 4 and 10) would indicate centrifugation to be a suitable method for artificially increasing the rate of sedimentation of a suspension. Again, it is not always possible to predict accurately the behaviour of such a system when stored under normal conditions from data obtained after this type of accelerated testing. The process of centrifugation may destroy the structure of a flocculated system that would remain intact under normal storage conditions. The sediment formed would become tightly packed and difficult to redisperse, whether or not the initial suspension was flocculated or deflocculated. This method may, however, give a useful indication of the relative stabilities of a series of trial products, particularly if used at speeds no faster than 200–300 rpm.

Rheological assessment

Although apparent viscosity measurements are also used as a tool to assess physical stability, the high shear rates involved may also destroy the structure of a suspension or emulsion. Very low rates of shear, using for example the Brookfield viscometer with Helipath stand, can give an indication of the change in the structure of the system after various storage times. For suspensions it may be possible to combine the results from sedimentation techniques with those from rheological assessments.

A measurement of the residual apparent viscosity, after breaking down the structure of the suspension, can be used as a routine quality control procedure after manufacture.

MANUFACTURE OF SUSPENSIONS

It is important to ensure initially that the powder to be suspended is in a suitably fine degree of subdivision in order to ensure adequate bioavailability, minimum sedimentation rate and impalpability. Suitable size reduction equipment and the relative merits of wet and dry milling are detailed in Chapter 11.

For the extemporaneous preparation of suspensions on a small scale, the powdered drug can be mixed with the suspending agent and some of the vehicle using a pestle and mortar. It may also be necessary, at this stage, to include a wetting agent to aid dispersion. Other soluble ingredients should then be dissolved in another portion of the vehicle,

mixed with the concentrated suspension and then made up to volume.

It is often preferable, particularly on a larger scale, to make a concentrated dispersion of the suspending agent first. This is best accomplished by adding the material slowly to the vehicle while mixing. Suitable mixers are described under Manufacture of emulsions, but can include either an impeller type of blender or a turbine mixer. This stage is important, as it is necessary to ensure that agglomerates of the suspending agent are fully broken up. If they are not, then the surface of each agglomerate may gel and cause the powder inside to remain non-wetted. Very intense shearing, however, can destroy the polymeric structure of the suspending agent, and it may be better to use milder shearing and then allow the dispersion to stand until full hydration has been achieved. This may be instantaneous or may, as with tragacanth, take several hours. If the suspending agent is blended with one of the water-soluble ingredients, such as sucrose, this will also aid dispersion.

The drug to be suspended is then added in the same way, along with the wetting agent. For very hydrophobic drugs, wetting may be facilitated by mixing under reduced pressure. This has the additional advantage of de-aerating the product and thus improving its appearance. Other ingredients should now be added, preferably dissolved in a portion of the vehicle, and the whole made up to volume if necessary. Finally, homogenization (see under Manufacture of emulsions) would ensure complete dispersion of the drug and the production of a smooth and elegant preparation.

It is also possible, though much less widely used, to suspend an insoluble drug by precipitating it from a solution. This can be accomplished either by double decomposition or, if it is a weak acid or a weak base, by altering the pH of its solution or by precipitating the drug from a water-miscible solvent on the addition of water. This method may be of use if the drug is required to be sterile but is degraded by heat or irradiation. A soluble form of the drug is dissolved in a suitable vehicle, sterilized by filtration and then precipitated to form a suspension.

In normal circumstances aqueous suspensions can be autoclaved, as long as the process does not adversely affect either physical or chemical stability.

MANUFACTURE OF EMULSIONS

It has already been explained that the smaller the globules of the disperse phase, the slower will be the rate of creaming in an emulsion. The size of these globules can also affect the viscosity of the product, and in general it has been found that the best emulsions with respect to physical stability and texture exhibit a mean globule diameter of between 0.5 and 2.5 μm. The choice of suitable equipment for the emulsification process depends mainly on the intensity of shearing required to produce this optimum particle size. Other considerations, however, include the volume and viscosity of the emulsion and the interfacial tension between the oil and the water. The presence of surfactants, which will reduce interfacial tension, will aid the process of emulsification as well as promoting emulsion stability.

In many cases simple blending of the oil and water phases with a suitable emulgent system may be sufficient to produce satisfactory emulsions. Further processing using a homogenizer can also be carried out to reduce globule size still further. The initial blending may be accomplished on a small scale by the use of a pestle and mortar or by using a mixer fitted with an impeller type of agitator, the size and type of which will depend primarily on the volume and viscosity of the emulsified product.

A more intense rate of shearing can be achieved using a turbine mixer such as the Silverson mixer–homogenizer. In this type of machine the short, vertical or angled rotor blades are enclosed within a stationary perforated ring and connected by a central rod to a motor. The liquids are therefore subjected to intense shearing, caused initially by the rotating blades, and then by the forced discharge through the perforated ring. Different models are available for a variety of batch sizes up to several thousand litres, and can include inline models.

The mixing vessel may also be fitted with baffles in order to modify the circulation of the liquid, and may be jacketed so that heating or cooling can be applied.

Homogenizers are often used after initial mixing to enable smaller globule sizes to be produced. They all work on the principle of forced discharge of the emulsion under pressure through fine interstices, formed by closely packed metal surfaces, in order to provide an intense shearing action.

If two immiscible liquids are subjected to ultrasonic vibrations, alternate regions of compression and rarefaction are produced. Cavities are then formed in the regions of rarefaction, which then collapse with considerable force causing emulsification. The required frequency of vibration is usually produced electrically, but mechanical methods are also available. Unfortunately this method of emulsification is limited to small-scale production.

Colloid mills are also suitable for the preparation of emulsions on a continuous basis. The intense shearing of the product between the rotor and the stator, which can be variably separated, will produce emulsions of very small globule size.

It is important to ensure that methods of manufacture developed on a laboratory scale can be easily extended to large-scale production, and without any change in the quality of the product.

During manufacture it is usual to add the disperse phase to the continuous phase during the initial mixing. The other ingredients are dissolved, prior to mixing, in the phase in which they are soluble. This is particularly important when making w/o emulsions. Oil-in-water emulsions, however, are sometimes made by the phase-inversion technique, in which the aqueous phase is slowly added to the oil phase during mixing. Initially a w/o emulsion is formed but, as further aqueous phase is added the emulsion inverts to form the intended product. This method often produces emulsions of very low mean droplet size.

Should any of the oily ingredients be of solid or semisolid consistency they must be melted before mixing. It is also essential that the aqueous phase be heated to the same temperature, to avoid premature solidification of the oil phase by the colder water on mixing but before emulsification has taken place. This also has the advantage of reducing the viscosity of the system, so enabling shear forces to be transmitted through the product more easily. Because of the increased kinetic motion of the emulgent molecules at the oil/water interface, however, it is necessary to continue stirring during the cooling process to avoid demulsification.

Volatile ingredients, including flavours and perfumes, are usually added after the emulsion has cooled. It must, however, be sufficiently fluid to enable adequate blending. Ingredients that may influence the physical stability of the emulsion, such as alcoholic solutions or electrolytes, require to be diluted as much as possible before adding slowly and with constant mixing.

RELEASE OF DRUGS FROM SUSPENSION AND EMULSION FORMULATIONS

Drug release from suspensions

After the oral administration of a suspension the drug, which is already in a wetted state, is presented to the gastrointestinal fluids in a finely divided form. Dissolution therefore occurs immediately. The rate of absorption of the drug into the bloodstream is therefore usually faster than for the same drug in a solid dosage form, but not as fast as that from a solution. The rate of release of a drug from a suspension is also dependent upon the viscosity of the product. The more viscous the preparation, the slower is likely to be the release of the drug. Care must therefore be taken to ensure that the physical characteristics of the suspension do not change on addition to an acid medium, if this should affect the rate of release of the drug.

Because the rate of release of an active agent from a suspension is usually slower than the release from solution, drugs are often formulated as suspensions for intramuscular, intraarticular or subcutaneous injection in order to prolong drug release. This is often termed depot therapy. Methylprednisolone, for example, which is available as the water-soluble sodium succinate salt, can be synthesized as the insoluble acetate ester. After intramuscular injection as a suspension, the rate of release is sufficiently slowed to maintain adequate blood levels for up to 14 days.

Release will occur even more slowly if the drug is suspended in an oil, which after injection will remain as a globule, so providing a minimal area of contact with tissue fluid.

Sustained release preparations formulated as suspensions for oral use are less common, but one example is the use of the Pennkinetic system. This involves the complexation of drugs such as hydrocortone and dextromethorphan with tiny ion exchange resin particles, which are then coated with ethylcellulose (Chang 1992). After ingestion the drug is slowly released by exchanging with ions present in the gastrointestinal tract. One of the main difficulties in the formulation of this type of product is to ensure that ions are not present in any of its ingredients.

Drug release from emulsions

The main commercial use of emulsions is for the oral, rectal and topical administration of oils and oil-soluble drugs. Lipid emulsions are also widely used for intravenous feeding, although the choice of emulgent is very limited and globule size must be kept below 4 μm diameter to avoid the formation of emboli. Quite often, however, the high surface area of dispersed oil globules will enhance the rate of absorption of lipophilic drugs.

The emulsion can also be used as a sustained-release dosage form. The intramuscular injection of certain water-soluble vaccines formulated as w/o emulsions can provide a slow release of the antigen

and result in a greater antibody response and hence a longer-lasting immunity. Other drugs have also been shown to have this effect, the rate of release being dependent mainly upon the oil/water partition coefficient of the drug and its rate of diffusion across the oil phase.

It is also possible to formulate multiple emulsion systems in which an aqueous phase is dispersed in oil droplets, which in turn are dispersed throughout another aqueous external phase, producing a water-in-oil-in-water (w/o/w) emulsion. These products can also be used for the prolonged release of drugs that are incorporated into the internal aqueous phase. These products have the advantage of exhibiting a lower viscosity than their w/o counterparts and hence are easier to inject.

Similarly, o/w/o emulsions can be formulated and are also under investigation as potential sustained-release bases.

Multiple emulsions, however, tend to be stable only for a relatively short time, although the use of polymers as alternatives to the traditional emulsifying agents may improve their physical stability (Florence and Whitehill 1982).

REFERENCES

Chang, R.K. (1992) Formulation approaches for sustained release oral suspensions. *Pharm. Tech.* **16**, 134–136.

Ciullo, P.A. (1981) Rheological properties of magnesium aluminium silicate/xanthan gum dispersions. *J. Soc. Cosmetic Chemists*, Sep/Oct, 32.

Florence, A.T. and Whitehill, D. (1982) Stabilisation of water/oil/water multiple emulsions by polymerisation of the aqueous phase. *J. Pharm. Pharmacol.*, **34**, 687–691.

Washington, C. (1990) The stability of intravenous fat emulsions in parenteral nutrition mixtures. *Int. J. Pharmaceutics*, **66**, 1–21.

BIBLIOGRAPHY

Florence, A.T. and Attwood, D. (1998) *Physicochemical Principles of Pharmacy*, 3rd Edn. Macmillan, Basingstoke, Chapter 7: Emulsions, suspensions and other dispersions, pp. 252–307; Chapter 8: Polymers and macromolecules, pp. 308–371.

24
Powders and granules

Malcolm Summers

CHAPTER CONTENTS

POWDERED AND GRANULATED PRODUCTS AS DOSAGE FORMS

The term 'powder' when used to describe a dosage form describes a formulation in which a drug powder has been mixed with other powdered excipients to produce the final product. The function of the added excipients depends upon the intended use of the product. Colouring, flavouring and sweetening agents, for example, may be added to powders for oral use.

Conventionally, the title 'powder' should be restricted to powder mixes for internal use, and alternative titles are used for other powdered formulations, e.g. dusting powders, which are for external use. The more descriptive title 'oral powder' to differentiate powders for internal use is recommended.

Granules which are used as a dosage form consist of powder particles that have been aggregated to form a larger particle, which is usually 2–4 mm in diameter. This is much larger than granules prepared as an intermediate for tablet manufacture. The processes involved in forming granules from powders are discussed in Chapter 25.

Powdered and granulated products are traditionally dispensed as:

- bulk powders or granules for internal use
- divided powders or granules (i.e. single preparations) for internal use
- dusting powders for external use.

Other preparations which are presented as powders or granules include:

- insufflations for administration to ear, nose or throat
- antibiotic syrups to be reconstituted before use
- powders for reconstitution into injections
- dry powder inhalers.

Advantages and disadvantages of powdered and granulated products

The advantages of this type of preparation are as follows.

1. Solid preparations are more chemically stable than liquid ones. The shelf-life of powders for antibiotic syrups, for example, is 2–3 years, but once they are reconstituted with water it is 1–2 weeks. The instability observed in liquid preparations is usually the primary reason for presenting some injections as powders to be reconstituted just prior to use.
2. Powders and granules are a convenient form in which to dispense drugs with a large dose. The dose of Compound Magnesium Trisilicate Oral Powder is 1–5 g, and although it is feasible to manufacture tablets to supply this dose it is often more acceptable to the patient to disperse a powder in water and swallow it as a draught.
3. Orally administered powders and granules of soluble medicaments have a faster dissolution rate than tablets or capsules, as these must first disintegrate before the drug dissolves. Drug absorption from such powdered or granulated preparations will therefore be faster than from the corresponding tablet or capsule, if the dissolution rate limits the rate of drug absorption.

The disadvantages of powders and granules are as follows.

1. Bulk powders or granules are far less convenient for the patient to carry than a small container of tablets or capsules, and are as inconvenient as liquid preparations, such as mixtures. Modern packaging methods for divided preparations, such as heat-sealable laminated sachets, mean that individual doses can be conveniently carried.
2. The masking of unpleasant tastes may be a problem with this type of preparation. A method of attempting this is by formulating the powder into a pleasantly tasting or taste-masked effervescent product, whereas tablets and capsules are a more common alternative for low-dose products.
3. Bulk powders or granules are not suitable for administering potent drugs with a low dose. This is because individual doses are extracted from the bulk using a 5 mL spoon. This method is subject to such variables as variation in spoon fill (e.g. 'level' or 'heaped' spoonfuls) and variation in the bulk density of different batches of a powder. It is therefore not an accurate method of measurement. Divided preparations have been used for more potent drugs, but tablets and capsules have largely replaced them for this purpose.
4. Powders and granules are not a suitable method for the administration of drugs which are inactivated in, or cause damage to, the stomach; these should be presented as enteric-coated tablets, for example.

DISPENSED PREPARATIONS

Bulk powders

The mixed ingredients are packed into a suitable bulk container, such as a wide-mouthed glass jar. Because of the disadvantages of this type of preparation the constituents are usually relatively non-toxic medicaments with a large dose, e.g. magnesium trisilicate and chalk, as present in Compound Magnesium Trisilicate Oral Powder. Relatively few proprietary examples exist, although many dietary/food supplements are packed in this way.

Divided powders

Divided powders are similar formulations to bulk powders but individual doses are separately wrapped.

Traditionally, single doses were wrapped in paper. This was unsatisfactory for most products, particularly if the ingredients were hygroscopic, volatile or deliquescent. Modern packaging materials of foil and plastic laminates have replaced such paper wrappings because they offer superior protective qualities and are amenable to use on high-speed packing machines.

Effervescent powders can now be packed in individual dose units because of the protective qualities of laminates. Such powders contain, for example, sodium bicarbonate and citric acid, which react and effervesce when the patient adds the powder to water to produce a draught. It is important to protect the powder from the ingress of moisture during manufacture and on subsequent storage to prevent the reaction occurring prematurely.

All powders and granules should be stored in a dry place to prevent deterioration due to ingress of moisture. Even if hydrolytic decomposition of susceptible ingredients does not occur, the particles will adhere and cake, producing an inelegant, often unusable product.

Bulk granules

One disadvantage of bulk powders is that, because of particle size differences, the ingredients may segregate (see Chapter 13), either on storage in the final container or in the hoppers of packaging machines. If this happens the product will be non-uniform and the patient will not receive the same dose of the ingredients on each occasion. This can be prevented by granulating the mixed powders.

Bulk granules therefore contain similar medicaments to powders, i.e. those with low-toxicity, high-dose drugs. Methylcellulose Granules, for example, are used as a bulk-forming laxative and have a dose of 1–4 g daily. Many proprietary preparations contain similar bulk-forming laxatives.

Divided granules

These are granulated products in which sufficient for one dose is individually wrapped. Effervescent granules can be formulated and presented in this manner. The comments on packaging materials discussed under Divided powders above are also equally pertinent to divided granules.

Dusting powders

Dusting powders contain ingredients used for therapeutic, prophylactic or lubricant purposes and are intended for external use.

Only sterile dusting powders should be applied to open wounds. Such preparations should be prepared using materials and methods designed to ensure sterility and to avoid the introduction of contaminants and the growth of micro-organisms.

Dusting powders for lubricant purposes or superficial skin conditions need not be sterile but they should be free from pathogenic organisms. As minerals such as talc and kaolin may be contaminated at source with spores of organisms causing tetanus and gangrene, these should be sterilized before they are incorporated into the product. Talc dusting powder is a sterile cutaneous powder containing starch and purified talc in which the talc is sterilized before incorporation with the starch, or the final product is subject to a suitable terminal sterilization procedure.

Dusting powders are normally dispensed in glass or metal containers with a perforated lid. The powder must flow well from such a container, so that they can be dusted over the affected area. The active ingredients must therefore be diluted with materials having reasonably good flow properties, e.g. purified talc or maize starch.

Hexachlorophane Dusting Powder contains an antibacterial agent and Talc Dusting Powder is used as a lubricant to prevent chafing. Proprietary products are available, usually for the treatment of bacterial or fungal infections, e.g. Canesten Powder (clotrimazole) is used as an antifungal agent and CX Powder (containing chlorhexidine acetate) is used as a general skin disinfectant.

Insufflations

Insufflations are medicated powders which are blown into regions such as the ear, nose and throat using an insufflator. The use of traditional insufflations declined because they were not very acceptable, being more inelegant and less convenient to apply than other topical preparations. A second problem was that if the powder contained a drug that had systemic activity it was difficult, with the conventional insulator, to ensure that the same dose was delivered on each occasion.

Some potent drugs are now presented in this way because they are rapidly absorbed when administered as a fine powder via the nose (see Chapter 32 for a detailed discussion of the nasal route of administration). To enhance convenience and ensure that a uniform dose is delivered on each occasion, devices have been developed to replace the traditional insufflator. Sufficient drug for one dose may be presented in a hard gelatin capsule diluted with an inert, soluble diluent such as lactose. The capsule is placed in the body of the insufflator and is broken when the device is assembled. The drug is inhaled by the patient as a fine powder.

Dry-powder inhalers

The use of dry-powder systems for pulmonary drug delivery is now extensive. This dosage form has developed into one of the most effective methods of delivering active ingredients to the lung for the treatment of asthma and chronic obstructive pulmonary disease. Its popularity is reflected in the number of commercial preparations available in a number of sophisticated and increasingly precise devices. Pulmonary delivery is discussed fully in Chapter 31, and the reader is referred there for further information.

PREPARATIONS REQUIRING FURTHER TREATMENT AT TIME OF DISPENSING

Oral antibiotic syrups

For patients who have difficulty taking capsules and tablets, e.g. young children, a liquid preparation of a drug offers a suitable alternative. However, many drugs, e.g. antibiotics, are physically or chemically unstable when formulated as a solution or suspension. The method used to overcome this instability problem is to manufacture the dry ingredients of the intended liquid preparation in a suitable container in the form of a powder or granules. When the pharmacist dispenses the product, a given quantity of water is added to reconstitute the solution or suspension. This enables sufficient time for warehouseing and distribution of the product and storage at the pharmacy without degradation. Once it is reconstituted, the patient is warned of the short shelf-life. A shelf-life of 1–2 weeks for the reconstituted syrup should not be a serious problem for the patient. Examples are Erythroped Suspension and Amoxicillin Oral Suspension.

Powders for injection

Injections of medicaments that are unstable in solution must be made immediately prior to use and are presented as sterile powders in ampoules. Sufficient diluent, e.g. sterile Water for Injections, is added from a second ampoule to produce the required drug concentration and the injection is used immediately. The powder may contain suitable excipients in addition to the drug, e.g. sufficient additive to produce an isotonic solution when the injection is reconstituted.

BIBLIOGRAPHY

British Pharmacopoeia (2000), HMSO, London.

Further information on proprietary products can be found in:

The Data Sheet Compendium, Datapharm Publications Ltd (published annually).

The Monthly Index of Medical Specialities (MIMS), Medical Publications Ltd.

25
Granulation

Malcolm Summers, Michael Aulton

CHAPTER CONTENTS

INTRODUCTION TO GRANULATION

Granulation is the process in which **primary powder particles** are made to adhere to form larger, multiparticle entities called **granules**. Pharmaceutical granules typically have a size range between 0.2 and 4.0 mm, depending on their subsequent use. In the majority of cases this will be in the production of tablets or capsules, when granules will be made as an intermediate product and have a typical size range between 0.2 and 0.5 mm, but larger granules are used as a dosage form in their own right (see Chapter 24).

Granulation normally commences after initial dry mixing of the necessary powdered ingredients so that a uniform distribution of each ingredient through the mix is achieved. After granulation the granules will either be packed (when used as a dosage form), or they may be mixed with other excipients prior to tablet compaction or capsule filling.

Reasons for granulation

The reasons why granulation is often necessary are as follows.

To prevent segregation of the constituents of the powder mix

Segregation (or demixing, see Chapter 13) is due primarily to differences in the size or density of the components of the mix, the smaller and/or denser particles concentrating at the base of a container with the larger and/or less dense ones above them. An ideal granulation will contain all the constituents of the mix in the correct proportion in each granule, and segregation of the ingredients will not occur (Fig. 25.1).

It is also important to control the particle size distribution of the granules because, although the individual components may not segregate, if there is a wide size distribution the granules themselves may segregate. If this occurs in the hoppers of sachet-filling machines, capsule-filling machines or tablet machines, products with large weight variations will result. This is because these machines fill by volume rather than weight, and if different regions in the hopper contain granules of different sizes (and hence bulk density), a given volume in each region will contain a different weight of granules. This will lead to an unacceptable distribution of the drug content within the batch of finished product, even though the drug is evenly distributed, weight per weight, through the granules.

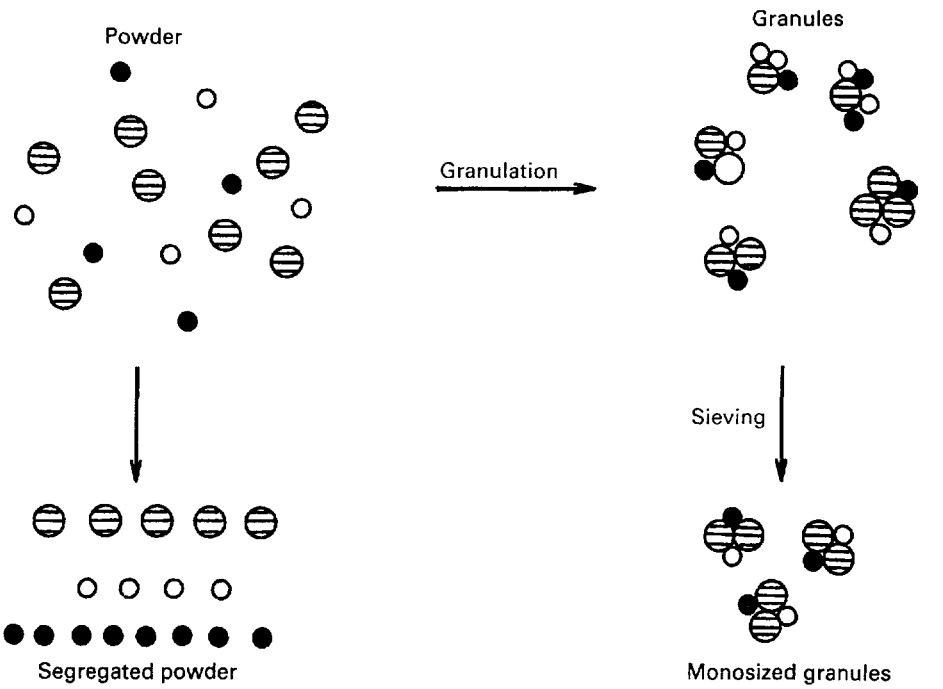


Fig. 25.1 Granulation to prevent powder segregation.

To improve the flow properties of the mix

Many powders, because of their small size, irregular shape or surface characteristics, are cohesive and do not flow well. Poor flow will often result in a wide weight variation within the final product owing to variable fill of tablet dies etc. Granules produced from such a cohesive system will be larger and more isodiametric, both factors contributing to improved flow properties.

To improve the compaction characteristics of the mix

Some powders are difficult to compact even if a readily compactable adhesive is included in the mix, but granules of the same formulation are often more easily compacted and produce stronger tablets. This is associated with the distribution of the adhesive within the granule and is a function of the method employed to produce the granule. Often solute migration (see Chapter 26) occurring during the postgranulation drying stage results in a binder-rich outer layer to the granules. This in turn leads to direct binder–binder bonding, which assists the consolidation of weakly bonding materials.

Other reasons

The above are the primary reasons for the granulation of pharmaceutical products, but there are other reasons that may necessitate the granulation of powdered material:

1. The granulation of toxic materials will reduce the hazard associated with the generation of toxic dust that may arise when handling powders. Suitable precautions must be taken to ensure that such dust is not a hazard during the granulation process. Thus granules should be non-friable and have a suitable mechanical strength.
2. Materials which are slightly hygroscopic may adhere and form a cake if stored as a powder. Granulation may reduce this hazard, as the granules will be able to absorb some moisture and yet retain their flowability because of their size.
3. Granules, being denser than the parent powder mix, occupy less volume per unit weight. They are therefore more convenient for storage or shipment.

Methods of granulation

Granulation methods can be divided into two types: **wet** methods, which use a liquid in the process, and ***dry*** methods in which no liquid is used.

In a suitable formulation a number of different excipients will be needed in addition to the drug. The common types used are diluents, to produce a unit dose weight of suitable size, and disintegrating agents, which are added to aid the break-up of the granule when it reaches a liquid medium, e.g. on ingestion by the patient. Adhesives in the form of a dry powder may also be added, particularly if dry granulation is employed. These ingredients will be mixed before granulation.

Dry granulation

In the dry methods of granulation the primary powder particles are aggregated under high pressure. There are two main processes. Either a large tablet (known as a '***slug***') is produced in a heavy-duty tabletting press (a process known as '***slugging***') or the powder is squeezed between two rollers to produce a sheet of material ('***roller compaction***'). In both cases these intermediate products are broken using a suitable milling technique to produce granular material, which is usually sieved to separate the desired size fraction. The unused fine material may be reworked to avoid waste. This dry method may be used for drugs that do not compress well after wet granulation, or those which are sensitive to moisture.

Wet granulation (involving wet massing)

Wet granulation involves the massing of a mix of dry ***primary powder particles*** using a ***granulating fluid***. The fluid contains a solvent which must be volatile so that it can be removed by drying, and be non-toxic. Typical liquids include water, ethanol and isopropanol, either alone or in combination. The granulation liquid may be used alone or, more usually, as a solvent containing a dissolved ***adhesive*** (also referred to as a ***binder*** or ***binding agent***) which is used to ensure particle adhesion once the granule is dry.

Water is commonly used for economical and ecological reasons. Its disadvantages as a solvent are that it may adversely affect drug stability, causing hydrolysis of susceptible products, and it needs a longer drying time than do organic solvents. This increases the length of the process and again may affect stability because of the extended exposure to heat. The primary advantage of water is that it is non-flammable, which means that expensive safety precautions such as the use of flameproof equipment need not be taken. Organic solvents are used when

water-sensitive drugs are processed, as an alternative to dry granulation, or when a rapid drying time is required.

In the traditional wet granulation method the wet mass is forced through a sieve to produce wet granules which are then dried. A subsequent screening stage breaks agglomerates of granules and removes the fine material, which can than be recycled. Variations of this traditional method depend on the equipment used, but the general principle of initial particle aggregation using a liquid remains in all of the processes.

Effect of granulation method on granule structure

The type and capacity of granulating mixers significantly influences the work input and time necessary to produce a cohesive mass, adequate liquid distribution and intragranular porosity of the granular mass. The method and conditions of granulation affect intergranular and intragranular pore structure by changing the degree of packing within the granules. It has been shown that precompressed granules, consisting of compressed drug and binder particles, are held together by simple bonding during compaction. Granules prepared by wet massing consist of intact drug particles held together in a sponge-like matrix of binder. Fluidized-bed granules are similar to those prepared by the wet massing process, but possess greater porosity and the granule surface is covered by a film of binding agent. With spray-dried systems the granules consist of spherical particles composed of an outer shell and an inner core of particles. Thus the properties of the granule are influenced by the manufacturing process.

GRANULATION MECHANISMS

Particle-bonding mechanisms

To form granules, bonds must be formed between powder particles so that they adhere and these bonds must be sufficiently strong to prevent breakdown of the granule to powder in subsequent handling operations.

There are five primary bonding mechanisms between particles:

1. Adhesion and cohesion forces in the immobile liquid films between individual primary powder particles;
2. Interfacial forces in mobile liquid films within the granules;
3. The formation of solid bridges after solvent evaporation;
4. Attractive forces between solid particles;
5. Mechanical interlocking.

Different types of mechanism were identified in each group and the ones discussed below are those that are relevant to pharmaceutical granulations.

Adhesion and cohesion forces in immobile films

If sufficient liquid is present in a powder to form a very thin, immobile layer, there will be an effective decrease in interparticulate distance and an increase in contact area between the particles. The bond strength between the particles will be increased because of this, as the van der Waals forces of attraction are proportional to the particle diameter and inversely proportional to the square of the distance of separation.

This situation will arise with adsorbed moisture and accounts for the cohesion of slightly damp powders. Although such films may be present as residual liquid after granules prepared by wet granulation have been dried, it is unlikely that they contribute significantly to the final granule strength. In dry granulation, however, the pressures used will increase the contact area between the adsorption layers and decrease the interparticulate distance, and this will contribute to the final granule strength.

Thin, immobile layers may also be formed by highly viscous solutions of adhesives, and so the bond strength will be greater than that produced by the mobile films discussed below. The use of starch mucilage in pharmaceutical granulations may produce this type of film.

Interfacial forces in mobile liquid films

During wet granulation liquid is added to the powder mix and will be distributed as films around and between the particles. Sufficient liquid is usually added to exceed that necessary for an immobile layer and to produce a mobile film. There are three states of water distribution between particles, which are illustrated in Figure 25.2.

At low moisture levels, termed the ***pendular state***, the particles are held together by lens-shaped rings of liquid. These cause adhesion because of the surface tension forces of the liquid/air interface and the hydrostatic suction pressure in the liquid bridge. When all the air has been displaced from between

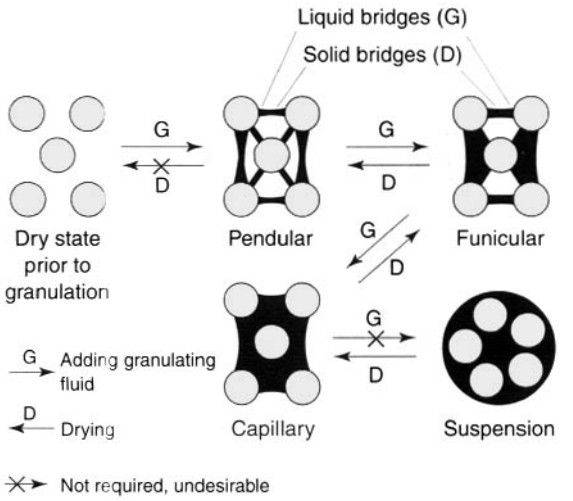


Fig. 25.2 Water distribution between particles of a granule during formation and drying.

the particles the ***capillary state*** is reached, and the particles are held by capillary suction at the liquid/air interface, which is now only at the granule surface. The ***funicular state*** represents an intermediate stage between the pendular and capillary states. Moist granule tensile strength increases about three times between the pendular and the capillary state.

It may appear that the state of the powder bed is dependent upon the total moisture content of the wetted powders, but the capillary state may also be reached by decreasing the separation of the particles. In the massing process during wet granulation, continued kneading/mixing of material originally in the pendular state will densify the wet mass, decreasing the pore volume occupied by air and eventually producing the funicular or capillary state without further liquid addition.

In addition to these three states, a further state, the droplet, is illustrated in Figure 25.2. This will be important in the process of granulation by spray-drying of a suspension. In this state, the strength of the droplet is dependent upon the surface tension of the liquid used.

These wet bridges are only temporary structures in wet granulation because the moist granules will be dried. They are, however, a prerequisite for the formation of solid bridges formed by adhesives present in the liquid, or by materials that dissolve in the granulating liquid.

Solid bridges

These can be formed by:

1. partial melting
2. hardening binders
3. crystallization of dissolved substances.

Partial melting Although not considered to be a predominant mechanism in pharmaceutical materials, it is possible that the pressures used in dry granulation methods may cause melting of low melting-point materials where the particles touch and high pressures are developed. When the pressure is relieved, crystallization will take place and bind the particles together.

Hardening binders This is the common mechanism in pharmaceutical wet granulations when an adhesive is included in the granulating solvent. The liquid will form liquid bridges, as discussed above, and the adhesive will harden or crystallize on drying to form solid bridges to bind the particles. Adhesives such as polyvinylpyrrolidone, the cellulose derivatives (such as carboxymethylcellulose) and pregelatinized starch function in this way.

Crystallization of dissolved substances The solvent used to mass the powder during wet granulation may partially dissolve one of the powdered ingredients. When the granules are dried, crystallization of this material will take place and the dissolved substance then acts as a hardening binder. Any material soluble in the granulating liquid will function in this manner, e.g. lactose incorporated into dry powders granulated with water.

The size of the crystals produced in the bridge will be influenced by the rate of drying of the granules: the slower the drying time, the larger the particle size. It is therefore important that the drug does not dissolve in the granulating liquid and recrystallize, because it may adversely affect the dissolution rate of the drug if crystals larger than that of the starting material are produced.

Attractive forces between solid particles

In the absence of liquids and solid bridges formed by binding agents, there are two types of attractive force that can operate between particles in pharmaceutical systems.

Electrostatic forces may be important in causing powder cohesion and the initial formation of agglomerates, e.g. during mixing. In general they do not contribute significantly to the final strength of the granule.

Van der Waals forces, however, are about four orders of magnitude greater than electrostatic forces and contribute significantly to the strength of granules produced by dry granulation. The magnitude of these forces will increase as the distance between

adjacent surfaces decreases, and in dry granulation this is achieved by using pressure to force the particles together.

Mechanisms of granule formation

In the dry methods, particle adhesion takes place because of applied pressure. A compact or sheet is produced which is larger than the granule size required, and therefore the required size can be attained by milling and sieving.

In wet granulation methods, liquid added to dry powders has to be distributed through the powder by the mechanical agitation created in the granulator. The particles adhere to each other because of liquid films, and further agitation and/or liquid addition causes more particles to adhere. The precise mechanism by which a dry powder is transformed into a bed of granules varies for each type of granulation equipment, but the mechanism discussed below serves as a useful broad generalization of the process.

The proposed granulation mechanism can be divided into three stages.

Nucleation

Granulation starts with particle–particle contact and adhesion due to liquid bridges. A number of particles will join to form the pendular state illustrated in Figure 25.2. Further agitation densifies the pendular bodies to form the capillary state, and these bodies act as nuclei for further granule growth.

Transition

Nuclei can grow in two possible ways: either single particles can be added to the nuclei by pendular bridges, or two or more nuclei may combine. The combined nuclei will be reshaped by the agitation of the bed.

This stage is characterized by the presence of a large number of small granules with a fairly wide size distribution. Providing that this distribution is not excessively large, this is a suitable end-point for granules used in capsule and tablet manufacture, as relatively small granules will produce a uniform tablet die or capsule fill. Larger granules may give rise to problems in small-diameter dies owing to bridging across the die and uneven fill.

Ball growth

Further granule growth produces large, spherical granules and the mean particle size of the granulating system will increase with time. If agitation is continued, granule coalescence will continue and produce an unusable, overmassed system, although this is dependent upon the amount of liquid added and the properties of the material being granulated.

Although ball growth produces granules that may be too large for pharmaceutical purposes, some degree of ball growth will occur in planetary mixers and it is an essential feature of some spheronizing equipment.

The four possible mechanisms of ball growth are illustrated in Figure 25.3.

Coalescence Two or more granules join to form a larger granule.

Breakage Granules break into fragments which adhere to other granules, forming a layer of material over the surviving granule.

Abrasion transfer Agitation of the granule bed leads to the attrition of material from granules. This abraded material adheres to other granules, increasing their size.

Layering When a second batch of powder mix is added to a bed of granules the powder will adhere to the granules, forming a layer over the surface and increasing the granule size. This mechanism is only relevant to the production of layered granules using spheronizing equipment.

There will be some degree of overlap between these stages and it will be very difficult to identify a given stage by inspection of the granulating system. For end-product uniformity it is desirable to finish every batch of a formulation at the same stage, and this may be a major problem in pharmaceutical production.

Using the slower processes, such as the planetary mixer, there is usually sufficient time to stop the process before overmassing occurs. With faster granulation equipment the duration of granulation can only be used as a control parameter when the formulation is such that granule growth is slow and takes place at a fairly uniform rate. In many cases, however, the transition from a non-granulated to an overmassed system is very rapid, and monitoring equipment is necessary to stop the granulation at a predetermined point, known as granulation end-point control.

PHARMACEUTICAL GRANULATION EQUIPMENT

Wet granulators

There are three main types of granulator used in the pharmaceutical industry for wet granulation.

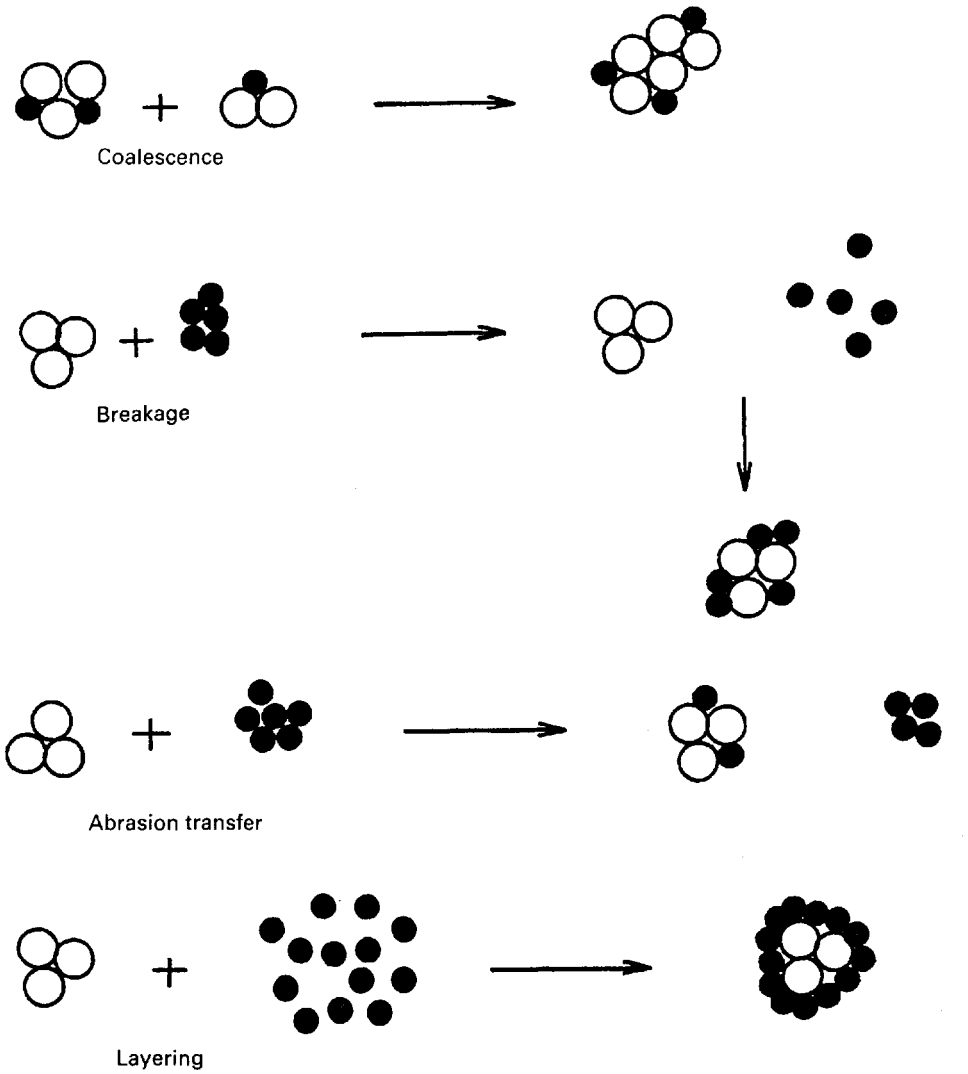


Fig. 25.3 Mechanisms of ball growth during granulation.

Shear granulators

In the traditional granulation process a planetary mixer is often used for wet massing of the powders, e.g. Hobart, Collette, Beken (Fig. 25.4). Powder mixing usually has to be performed as a separate operation using suitable mixing equipment. With some formulations, such as those containing two or three ingredients in approximately equal quantities, however, it may be possible to achieve a suitable mix in the planetary mixer without a separate stage.

The mixed powders are fed into the bowl of the planetary mixer and granulating liquid is added as the paddle of the mixer agitates the powders. The planetary action of the blade when mixing is similar to that of a household mixer.

The moist mass has then to be transferred to a granulator, such as an oscillating granulator (Fig. 25.5). The rotor bars of the granulator oscillate and force the moist mass through the sieve screen, the size of which determines the granule size. The mass should be sufficiently moist to form discrete granules when sieved. If excess liquid is added, strings of material will be formed and if the mix is too dry the mass will be sieved to powder and granules will not be formed.

The granules can be collected on trays and transferred to a drying oven, although tray drying suffers from three major disadvantages:

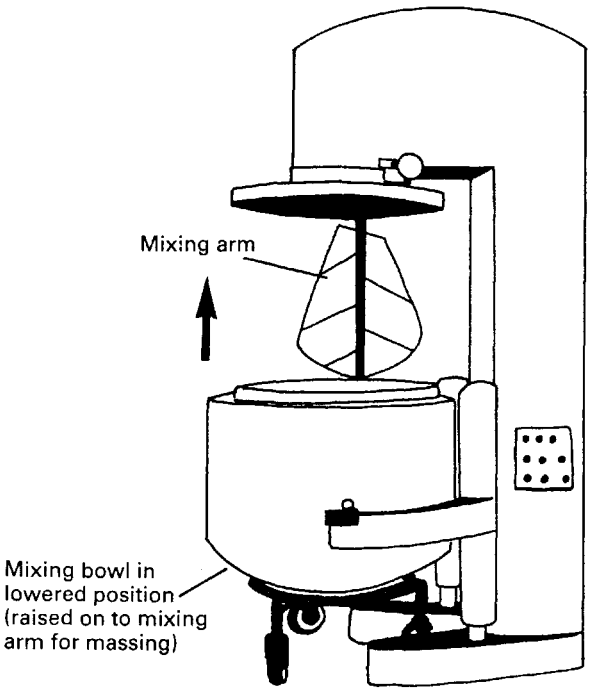


Fig. 25.4 Planetary mixer for wet massing.

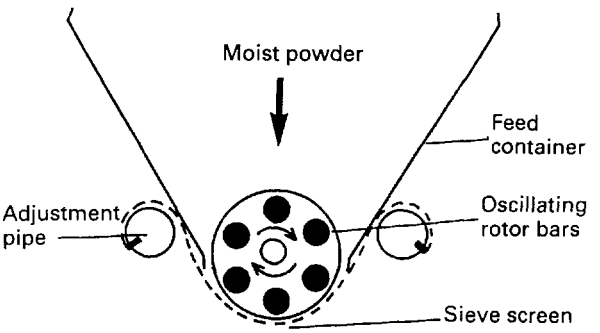


Fig. 25.5 Oscillating granulator.

1. The drying time is long.
2. Dissolved material can migrate to the upper surface of the bed of granules, as the solvent is only removed from the upper surface of the bed on the tray.
3. Granules may aggregate owing to bridge formation at the points of contact of the granules.

To deaggregate the granules and remix them, a sieving stage is necessary after drying.

An alternative method is to dry the granules using a fluidized-bed drier. This is quicker and, as it keeps the individual granules separated during drying, it reduces the problems of aggregation and intergranular solute migration, thereby reducing the need for a sieving stage after drying.

The disadvantages of this traditional granulation process are its long duration, the need for several pieces of equipment, and the high material losses that can be incurred because of the transfer stages. Advantages are that the process is not very sensitive to changes in the characteristics of the granule ingredients (e.g. surface area variations in different batches of an excipient), and the end-point of the massing process can often be determined by inspection.

High-speed mixer/granulators

This type of granulator (e.g. Diosna, Fielder) is used extensively in pharmaceutics. The machines have a stainless steel mixing bowl containing a three-bladed main impeller, which revolves in the horizontal plane, and a three-bladed auxiliary chopper (breaker blade) which revolves either in the vertical or the horizontal plane (Fig. 25.6).

The unmixed dry powders are placed in the bowl and mixed by the rotating impeller for a few minutes. Granulating liquid is then added via a port in the lid of the granulator while the impeller is turning. The granulating fluid is mixed into the powders by the impeller. The chopper is usually switched on when the moist mass is formed, as its function is to break up the wet mass to produce a bed of granular material. Once a satisfactory granule has been produced, the granular product is discharged, passing through a wire mesh which breaks up any large aggregates, into the bowl of a fluidized-bed drier.

The advantage of the process is that mixing, massing and granulation are all performed within a few minutes in the same piece of equipment. The process needs to be controlled with care as the granulation progresses so rapidly that a usable granule can be transformed very quickly into an unusable, overmassed system. Thus it is often necessary to use a suitable monitoring system to indicate the end of the granulation process, i.e. when a granule of the desired properties has been attained. The process is also sensitive to variations in raw materials, but this may be minimized by using a suitable end-point monitor.

A variation of the Diosna/Fielder type of design is the Collette–Gral mixer (Fig. 25.7). This is based on the bowl and overhead drive of the planetary mixer, but the single paddle is replaced by two mixing shafts. One of these carries three blades, which rotate in the horizontal plane at the base of the bowl, and the

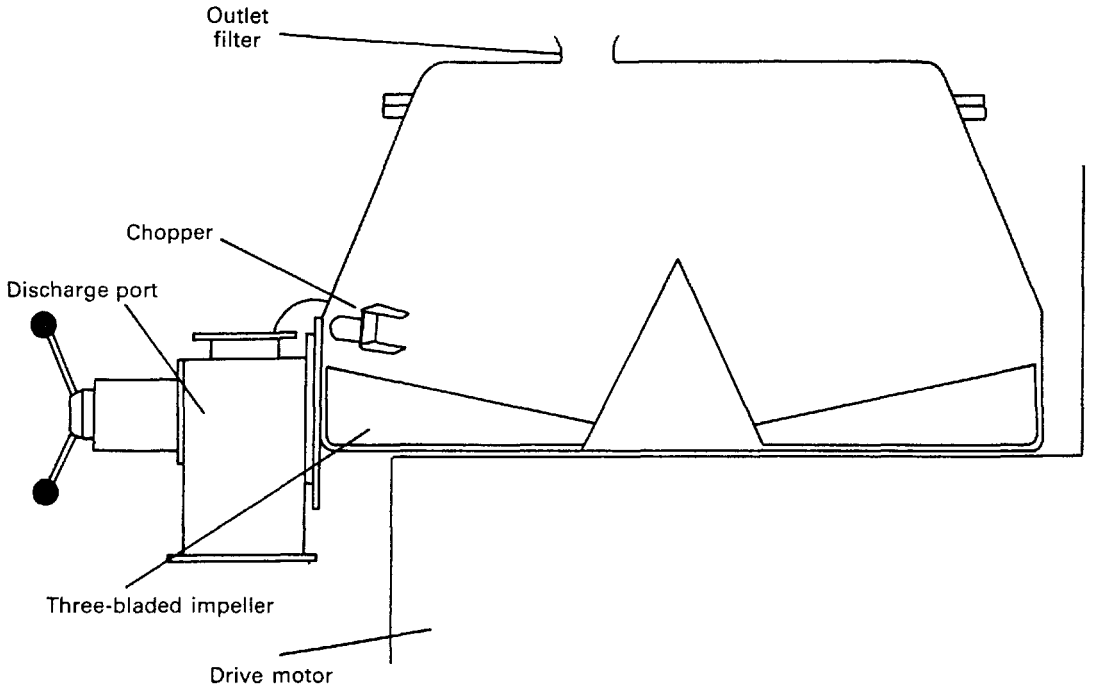


Fig. 25.6 High-speed mixer/granulator.

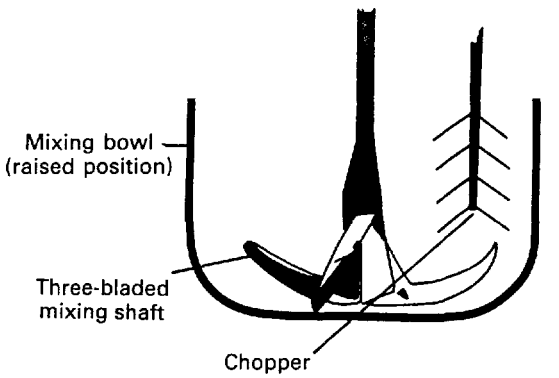


Fig. 25.7 Collette–Gral granulator: mixing shafts and bowl.

second carries smaller blades which act as the chopper and rotate in the horizontal plane in the upper regions of the granulating mass. Thus the operation principle is similar.

Fluidized-bed granulators

Fluidized-bed granulators (e.g. Aeromatic, Glatt) have a similar design and operation to fluidized-bed driers, i.e. the powder particles are fluidized in a stream of air, but in addition granulation fluid is sprayed from a nozzle on to the bed of powders (Fig. 25.8).

Heated and filtered air is blown or sucked through the bed of unmixed powders to fluidize the particles and mix the powders; fluidization is actually a very efficient mixing process. Granulating fluid is pumped from a reservoir through a spray nozzle positioned over the bed of particles. The fluid causes the primary powder particles to adhere when the droplets and powders collide. Escape of material from the granulation chamber is prevented by exhaust filters, which are periodically agitated to reintroduce the collected material into the fluidized bed. Sufficient liquid is sprayed to produce granules of the required size, at which point the spray is turned off but the fluidizing air continued. The wet granules are then dried in the heated fluidizing airstream.

Advantages of fluidized-bed granulation Fluidized-bed granulation has many advantages over conventional wet massing. All the granulation processes, which require separate equipment in the conventional method, are performed in one unit, saving labour costs, transfer losses and time. Another advantage is that the process can be automated once the conditions affecting the granulation have been optimized.

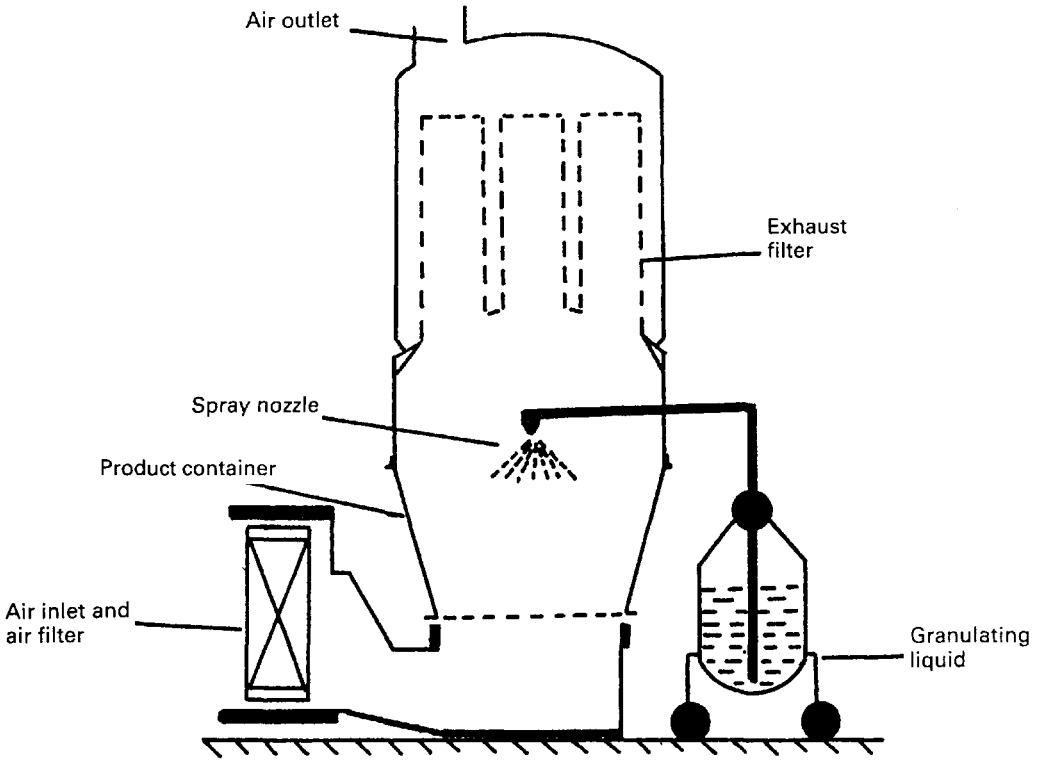


Fig. 25.8 Fluidized-bed granulator.

Disadvantages of fluidized-bed granulation On the downside, the equipment is initially expensive and optimization of process (and product) parameters affecting granulation needs extensive development work, not only during initial formulation work but also during scale-up from development to production. Similar development work for the traditional process and that using high-speed granulators is not as extensive.

This long and very product-specific development process has proved to be a serious problem with fluidized-bed granulation in the pharmaceutical industry. There are numerous apparatus, process and product parameters that affect the quality of the final granule. These are listed in Table 25.1. The extent of this list, coupled with the fact that each formulation presents its own individual development problems,

Table 25.1 Apparatus, process and product variables influencing fluidized-bed granulation

Apparatus parameters	Process parameters	Product parameters
Air distribution plate	Bed load	Type of binder
Shape of granulator body	Fluidizing air flow rate	Quantity of binder
Nozzle height	Fluidizing air temperature	Binder solvent
Positive or negative pressure operation	Fluidizing air humidity	Concentration of granulating solution
Scale-up	Atomization	Temperature of granulation solution
	Nozzle type	Starting Materials
	Spray angle	Fluidization
	Spraying regime	Powder hydrophobicity
	Liquid flow rate	
	Atomizing air flow rate	
	Atomizing air pressure	
	Droplet size	

has led to fluidized-bed granulation not reaching its full potential in pharmaceutical production. This is exacerbated by the reality that most pharmaceutical companies have a wide range of products made at relatively small batch sizes, unlike other industries (fertilizers, herbicides, foodstuffs) where fluidized-bed granulation is used successfully and extensively.

Spray-driers

These differ from the method discussed above in that a dry, granular product is made from a solution or a suspension rather than initially dry primary powder particles. The solution or suspension may be of drug alone, a single excipient or a complete formulation.

The process of spray-drying is discussed fully in Chapter 26. The resultant granules are free-flowing hollow spheres and the distribution of the binder in such granules (at the periphery following solute migration during drying) results in good compaction properties.

This process can be used to make tablet granules, although it is probably economically justified for this purpose only when suitable granules cannot be produced by the other methods. Spray-drying can convert hard elastic materials into more ductile ones. Spray-dried lactose is the classic example, and its advantages over α-lactose monohydrate crystals when compacted are discussed in Chapter 27.

The primary advantages of the process are the short drying time and the minimal exposure of the product to heat owing to the short residence time in the drying chamber. This means that little deterioration of heat-sensitive materials takes place, and it may be the only process suitable for this type of product.

Spheronizers/pelletizers

For some applications it may be desirable to have a dense, spherical pellet of the type difficult to produce with the equipment above. Such pellets are used for controlled drug release products following coating with a suitable polymer coat and filling into hard gelatin capsules. Capsule filling with a mixture of coated and non-coated drug-containing pellets would give some degree of programmed drug release after the capsule shell dissolves.

A commonly used process involves the separate processes of wet massing, followed by extrusion of this wet mass into rod-shaped granules and subsequent spheronization of these granules. Because this process is used so frequently to produce modified-release multiparticulates, it will be discussed in some detail.

Extrusion/spheronization

Extrusion/spheronization is a multistep process used to make uniformly sized spherical particles. It is used primarily to produce multiparticulates for controlled drug release applications. The major advantage over other methods of producing drug-loaded spheres or pellets is the ability to incorporate high levels of active ingredients without producing excessively large particles (i.e. minimal excipients are necessary).

The main steps of the process are:

1. ***Dry mixing of ingredients*** to achieve a homogenous powder dispersion;
2. ***Wet massing*** to produce a sufficiently plastic wet mass;
3. ***Extrusion*** to form rod-shaped particles of uniform diameter;
4. ***Spheronization*** to round off these rods into spherical particles;
5. ***Drying*** to achieve the desired final moisture content;
6. ***Screening*** (optional) to achieve the desired narrow size distribution.

Applications of extrusion/spheronization

Potential applications are many, but relate mainly to controlled drug release and improved processing.

Controlled drug release Both immediate-release and controlled-release pellets can be formed. In turn, these pellets can either be filled into hard gelatin capsule shells or compacted into tablets to form unit dosage forms. Pellets can contain two or more ingredients in the same individual unit, or incompatible ingredients can be manufactured in separate pellets.

Pellets can be coated in sub-batches to give, say, rapid-, intermediate- and slow-release pellets in the same capsule shell. Dense multiparticulates disperse evenly within the GI tract and have less variable gastric emptying and intestinal transit times than do single units, such as coated monolithic tablets.

Processing The process of extrusion/spheronization can be used to increase the bulk density, improve flow properties and reduce the problems of dust usually encountered with low-density, finely divided active and excipient powders.

Extrusion/spheronization is a more labour-intensive process than other forms of granulation and should therefore only be considered when other methods are either not satisfactory for that particular formulation or are inappropriate (i.e. when spheres are required).

Desirable properties of pellets

Uncoated pellets:

- Uniform spherical shape
- Uniform size
- Good flow properties
- Reproducible packing (into hard gelatin capsules)
- High strength
- Low friability
- Low dust
- Smooth surface
- Ease of coating.

Once coated:

- Maintain all of the above properties
- Have desired drug-release characteristics.

The process

Dry mixing of ingredients This uses normal powder-mixing equipment.

Wet massing This stage also employs normal equipment and processes as used in wet granulation. There are two major differences in the granulation step compared with granulation for compaction:

1. The amount of granulation fluid
2. The importance of achieving a uniform dispersion of fluid.

The amount of fluid needed to achieve spheres of uniform size and sphericity is likely to be greater than that for a similar tablet granulation. Poor liquid dispersion will produce a poor-quality product.

Extrusion Extrusion produces rod-shaped particles of uniform diameter from the wet mass. The wet mass is forced through dies and shaped into small cylindrical particles with uniform diameter. The extrudate particles break at similar lengths under their own weight. Thus the extrudate must have enough plasticity to deform, but not so much that the extruded particles adhere to other particles when collected or rolled in the spheronizer.

There are many designs of extruder, but generally they can be divided into three classes, based on their feed mechanism:

- Screw-feed extruders (axial or end-plate, dome and radial)
- Gravity-feed extruders (cylinder roll, gear roll, radial)
- Piston-feed extruders (ram).

The first two categories (Fig. 25.9) are used for both development and production, but the latter is only used for experimental development work as it is easy to add instrumentation.

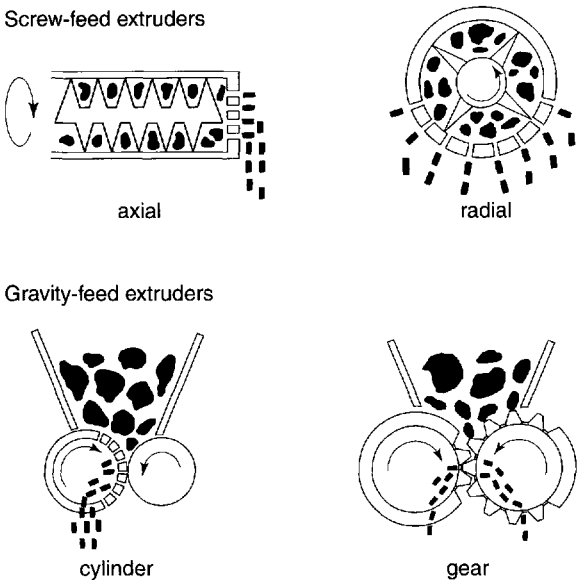


Fig. 25.9 Schematic representation of production extruders.

The primary extrusion process variables are:

- the feed rate of the wet mass
- the diameter of the die
- the length of the die
- the water content of the wet mass.

The properties of the extrudate, and hence the resulting spheres, are very dependent on the plasticity and cohesiveness of the wet mass. In general, an extrudable wet mass needs to be wetter than that appropriate for conventional granulation by wet massing.

Spheronization The function of the fourth step in the process (i.e. spheronization) is to round off the rods produced by extrusion into spherical particles.

This is carried out in a relatively simple piece of apparatus (Fig. 25.10). The working part consists of a bowl with fixed side walls and a rapidly rotating bottom plate or disc. The rounding of the extrudate into spheres is dependent on frictional forces generated by particle–particle and particle–equipment collisions. The bottom disc has a grooved surface to increase these forces. Two geometric patterns are generally used:

- A cross-hatched pattern with grooves running at right-angles to one another
- A radial pattern with grooves running radially from the centre of the disc.

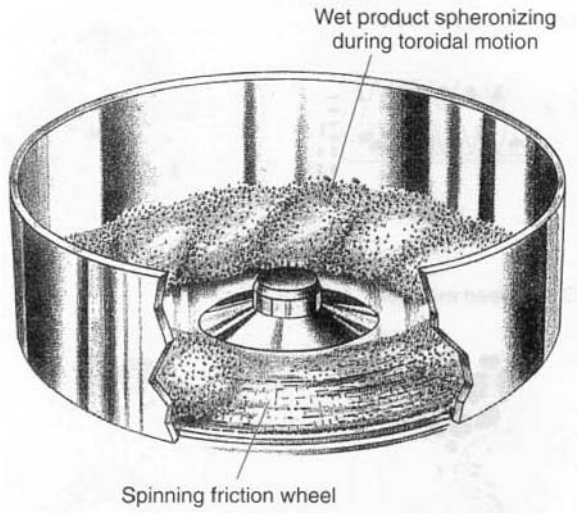


Fig. 25.10 A spheronizer showing the characteristic toroidal (rope-like) movement of the forming pellets in the spheronizer bowl during operation.

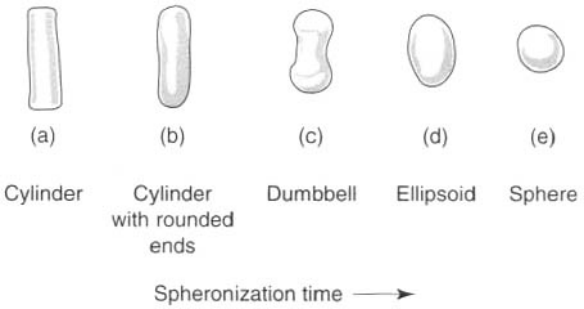


Fig. 25.11 Representation of a mechanism of spheronization. The diagram shows a transition from cylindrical particles (a) into cylindrical particles with rounded edges (b), then dumbbells (c), to ellipsoids (d) and finally spheres (e).

The transition from rods to spheres during spheronization occurs in various stages. These are best described by examining the following diagrams (Fig. 25.11).

If the mass is too dry spheres will not be formed: the rods will only transform as far as dumbbells.

Drying A drying stage is required in order to achieve the desired moisture content. This is often the final step in the process. The pellets can be dried in any drier that can be used for conventional wet granulations, including tray dryers and fluidized-bed driers. Both are used successfully for extrusion/spheronization. If solute migration (Chapter 26) occurs during drying of the wet spheres, this may result in:

- an increased initial rate of dissolution
- stronger pellets
- modified surfaces which might reduce the adhesion of any added film coats.

Screening Screening may be necessary to achieve the desired narrow size distribution. Normal sieves are used. If all the previous stages are performed efficiently and with careful development of process and formulation conditions, this step may not be necessary.

Formulation variables

The composition of the wet mass is critical in determining the properties of the particles produced. During the granulation step a wet mass is produced which must be plastic, deform when extruded, and break off to form uniformly sized cylindrical particles which are easily deformed into spherical particles. Thus the process has a complex set of requirements that are strongly influenced by the ingredients of the pellet formulation.

Summary

Extrusion/spheronization is a versatile process capable of producing spherical granules with very useful properties. Because it is more labour-intensive than more common wet massing techniques its use should be limited to those applications where a sphere is required and other granulation techniques are unsuitable.

The most common application of the process is to produce spherical pellets for controlled drug release.

Care must be taken to understand the required properties of the pellets and the manner in which the process and formulation influence the ability to achieve these aims.

Rotor granulation

This process allows the direct manufacture of spheres from dry powder. In the Freund granulator, the powder mix is added to the bowl and wetted with granulating liquid from a spray (Fig. 25.12). The baseplate rotates at high speed and centrifugal force keeps the moist mass at the edges of the rotor. Here, the velocity difference between the rotor and the static walls, combined with the upward flow of air around the rotor plate, causes the mass to move in a toroidal motion, resulting in the formation of discrete spherical pellets. These spheres (actually, of course, wet granules) are dried by the heated inlet air

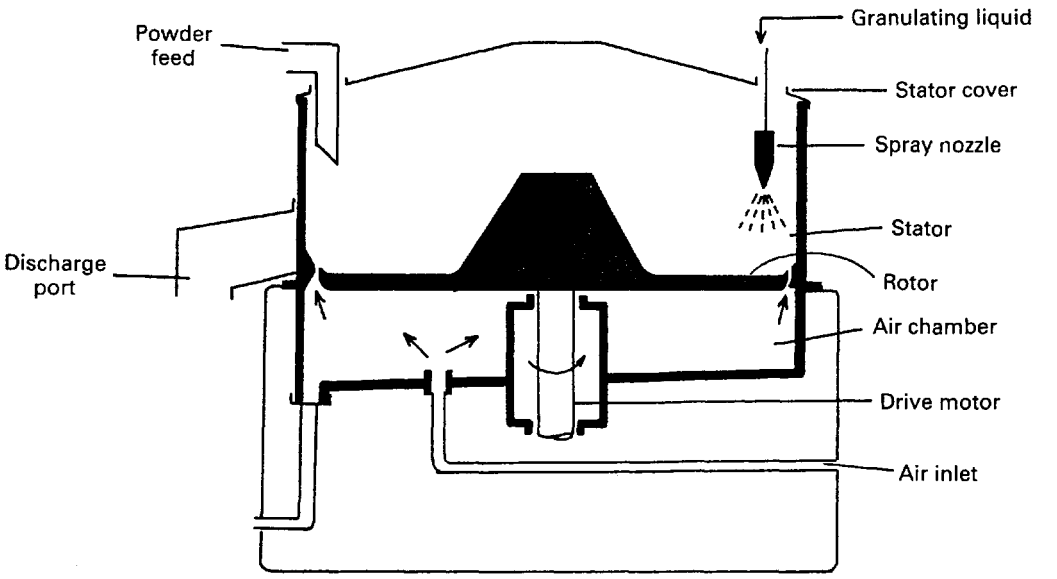


Fig. 25.12 Freund granulator.

from the air chamber, which also acts as a positive-pressure seal during granulation.

Using this technique it is possible to continue the process and coat the pellets by subsequently spraying coating solution on to the rotating dried pellets. In addition, layered pellets can be produced by using uncoated pellets as nuclei in a second granulation with a powder mix of a second ingredient or ingredients.

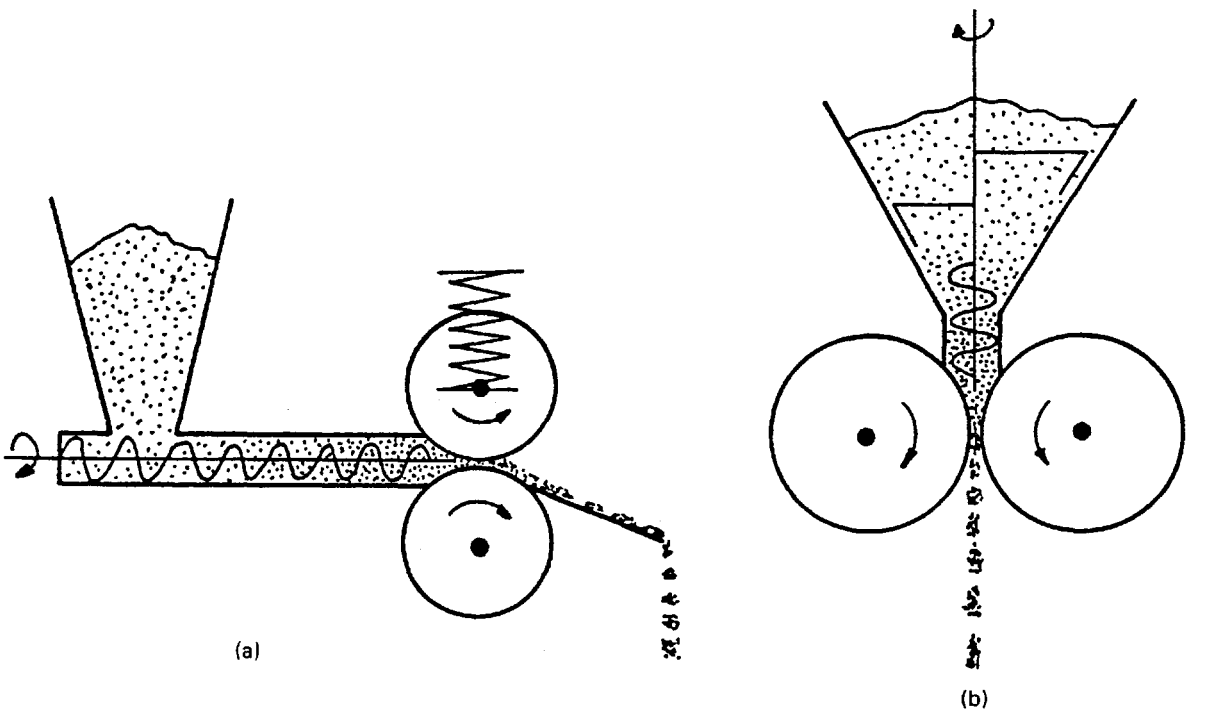


Fig. 25.13 Roller compaction: (a) Alexanderwerk and (b) Hutt types.

Dry granulators

Dry granulation converts primary powder particles into granules using the application of pressure without the intermediate use of a liquid. It therefore avoids heat–temperature combinations that might cause degradation of the product.

Two pieces of equipment are necessary for dry granulation: first, a machine for compressing the dry powders into compacts or flakes, and secondly a mill for breaking up these intermediate products into granules.

Sluggers

The dry powders can be compressed using a conventional tablet machine or, more usually, a large heavy-duty rotary press can be used. This process is often known as 'slugging', the compact made in the process (typically 25 mm diameter by about 10–15 mm thick) being termed a 'slug'. A hammer mill is suitable for breaking the compacts.

Roller compactors

Roller compaction is an alternative gentler method, the powder mix being squeezed between two rollers to form a compressed sheet (Fig. 25.13). The sheet normally is weak and brittle and breaks immediately into flakes. These flakes need gentler treatment to break them into granules, and this can usually be achieved by screening alone.

BIBLIOGRAPHY

There are a large number of published papers in the field of pharmaceutical granulation and only a limited number are listed below.

Baert, L. (1992) Correlation of extrusion forces, raw materials and sphere characteristics. *J. Pharm. Pharmacol.*, **44**, 676–678.

Das, S. and Jarowski, C.I. (1979) Effect of granulating method on particle size distribution of granules and disintegrated tablets. *Drug Dev. Ind. Pharm.*, **5**, 479.

Faure, A., Grimsey, I.M., Rowe, R.C. York, P. and Cliff, M.J. (1999) Applicability of a scale-up methodology for the wet granulation process in Collette Gral high shear mixer granulators. *Eur. J. Pharm. Sci.*, **8**, 85–93.

Gandhi, R., Kaul, C.L. and Panchagnula, R. (1999). Extrusion and spheronization in the development of oral controlled-release dosage forms. *Pharm. Sci. Tech. Today*, **2(4)**, 160–81.

Gergely, G. (1981) Granulation – a new approach. *Mfg Chem. Aerosol News*, **52**, 43.

Landin, M., York, P., Cliff, M.J. and Rowe, R.C. (1999) Scale up of a pharmaceutical granulation in planetary mixers. *Pharm. Dev. Tech.*, **4(2)**, 145–150.

Law, M.F.L. et al (1997). Comparison of two commercial brands of microcrystalline cellulose for extrusion–spheronization. *J. Microencap.*, **14(6)**, 713–723.

Lindberg, N.-O. and Leander, L. (1982) Instrumentation of a Kenwood major domestic-type mixer for studies of granulation. *Drug Dev. Ind. Pharm.*, **8**, 775.

Maejima, T. et al (1998) Application of tumbling melt granulation (TMG) method to prepare controlled-release fine granules. *Chem. Pharm. Bull*, **46(3)**, 534–536.

Nurnberg, E. and Wunderlich, J. (1998). Manufacturing pellets by extrusion and spheronization (Part I). *Pharm. Technol. Eur.*, **11(2)**, 41–47; Part II **11(3)**.

Ogawa, S. et al (1994) A new attempt to solve the scale-up problem for granulation using response surface methodology. *J. Pharm. Sci.*, **83(3)**, 439–443.

Parikh, D.M. (1997). *Handbook of Pharmaceutical Granulation Technology*, Marcel Dekker, New York.

Rubino, O.R. (1999) Fluid-bed technology; overview and criteria for process selection. *Pharm. Tech.*, **(6)**, 104–113.

Shah, R.D. et al (1995) Physico-mechanical characterization of the extrusion-spheronization process. *Pharm. Res.*, **12(4)**, 496–507.

Thoma, K. and Ziegler, I. (1998) Investigations on the influence of the type of extruder for pelletization by extrusion–spheronization I. Extrusion behavior of formulations. *Drug Dev. Ind. Pharm.*, **24(5)**, 401–411; II Sphere characteristics. *Drug Dev. Ind. Pharm.*, **24(5)**, 413–422.

Wørts, O. (1998) Wet granulation – fluidized bed and high shear techniques compared. *Pharm. Techn. Europe*, **10(11)**, 27–30.

Zhang, F. and McGinity, J.M. (1999). Properties of sustained-release tablets prepared by hot-melt etrusion. *Pharm. Dev. and Tech.*, **4(2)**, 241–250.

26
Drying

Michael Aulton

CHAPTER CONTENTS

INTRODUCTION

Drying is an important operation in primary pharmaceutical manufacture (i.e. the synthesis of actives) as it is usually the last stage of manufacturing before packaging, and it is important that the residual moisture is rendered low enough to prevent product deterioration during storage and ensure free-flowing properties during use. It is equally important (and probably encountered more frequently) in secondary (dosage form) manufacture following the commonly performed operation of wet granulation (Chapter 25) during the preparation of granules prior to tablet compaction. Hence, stability, flow properties and compactability are all influenced by residual moisture (see Chapter 27).

This chapter is concerned with drying to the solid state, starting with either a wet solid or a solution or suspension of the materials that will form the final dry product. Formerly solutions or suspensions were first concentrated by evaporation of much of the liquid before drying of the intermediate paste so formed. Equipment such as the spray drier (see later) are now capable of producing a dry product from a solution or suspension in one operation.

Most pharmaceutical materials are not completely free from moisture ('bone dry') but contain some residual water, which may vary with the temperature and humidity of the ambient air to which they are exposed. This is discussed in more detail later.

Some wet crystallized product may be freed from much of its residual moisture by initial treatment in a 'dewatering' centrifuge, which is similar in construction to a domestic spindrier. For the purpose of this chapter, however, drying is defined as the removal of all or most of the liquid by supplying latent heat to cause thermal vaporization, i.e. a liquid is converted into a vapour. In the majority of cases the 'liquid' will be water, but volatile solvents such as isopropanol may also need to be removed in a drying process. The physical principles are similar regardless of the nature of the liquid, although volatile solvents are normally recovered by condensation (rather than being vented into the atmosphere). The toxicity and flammability of organic solvents pose safety considerations.

THE DRYING OF WET SOLIDS

An understanding of this operation requires some preliminary explanation of the following important terms. These will be defined and explained in the context of water (the most commonly used pharmaceutical solvent), but the explanations and concepts are equally applicable to other relevant liquids (e.g. ethanol, isopropanol etc).

Moisture content of wet solids

The moisture content of a wet solid is expressed as kilograms of moisture associated with 1 kg of the moisture-free or 'bone-dry' solid. Thus a moisture content of 0.4 means that 0.4 kg of removable water is present per kg of the 'bone-dry' solid which will remain after complete drying. It is sometimes calculated as a percentage moisture content.

Total moisture content

This is the total amount of liquid associated with a wet solid. In the context of drying, not all of it can be easily removed by the simple evaporative processes employed by most pharmaceutical driers. The easily removable water is known as the ***free moisture content***, and the moisture which is more difficult to remove in practice is the ***equilibrium moisture content*** (see below). The easily removable water is known as ***unbound water***.

Unbound water This water exists as a liquid and exerts its full vapour pressure; it can be removed readily by evaporation. During a drying process this water is easily lost but the resulting solid is not completely free from water molecules; this is known as ***air dry***.

Equilibrium moisture content

Evaporative drying processes will not remove all the possible moisture present in a wet product because the solid equilibrates with the moisture present in the air. The moisture content present in a solid under steady-state ambient conditions is termed the equilibrium moisture content. Its value changes with temperature, humidity and the nature of the solid.

Bound water Part of the moisture present in a wet solid may be adsorbed on surfaces of the solid or be adsorbed within its structure to such an extent to prevent it from developing its full vapour pressure and from being easily removed by evaporation. Such moisture is described as 'bound' and is more difficult to remove than unbound water. The adsorbed water is attached to the surface of the solid as individual water molecules, which may form a mono- (or bi-) layer on the solid surface. Absorbed water exists as a liquid but is trapped in capillaries within the solid by

surface tension. As such it cannot exert its full vapour pressure and is not easily lost by evaporation.

Relative humidity (RH) of air

Air at a given temperature is capable of taking up water vapour until it is saturated (at 100% RH). It is a simple solution of water in air that follows the rules of most solutions – such as increased solubility with increasing temperature, a maximum solubility at a particular temperature (saturation) and precipitation of the solute on cooling (condensation, rain). If the temperature is raised then the air will be able to take up more moisture and the relative humidity falls.

This is because the percentage RH may be defined as:

$$\frac{\text{Vapour pressure of water vapour in the air}}{\text{Vapour pressure of water vapour in air saturated at the same temperature}} \times 100$$

This is ***approximately*** equal to the ***percentage saturation***, which is the ratio:

$$\frac{\text{Mass of vapour present per kg of dry air}}{\text{Mass of vapour required to saturate 1 kg of air at the same temperature}} \times 100$$

This relationship shows that the relative humidity of air is dependent not only on the amount of moisture in the air, but also on its temperature, as the amount of water required to saturate air is itself dependent on temperature.

It should be noted that in convective drying, where warm air is passed over the surface of a wet solid, the relative humidity may rise during the drying process as a result of two separate factors:

1. Uptake of evaporated water vapour from the wet solid;
2. The cooling of the supply air as it transfers heat to the wet solid (evaporative cooling).

If the cooling is excessive the temperature of the air may fall to a value known as the ***dew point***, when liquid water will condense and be deposited.

Wet-bulb and dry-bulb temperature

If two similar thermometers are set up, one with its bulb kept moist by a wet cotton wick immersed in a water reservoir, the wet-bulb thermometer will register a lower temperature than its dry-bulb neighbour (Fig. 26.1). This is due to the ***evaporative cooling*** as the latent heat of evaporation (Chapter 38) is taken from the sensible heat of the water surrounding the

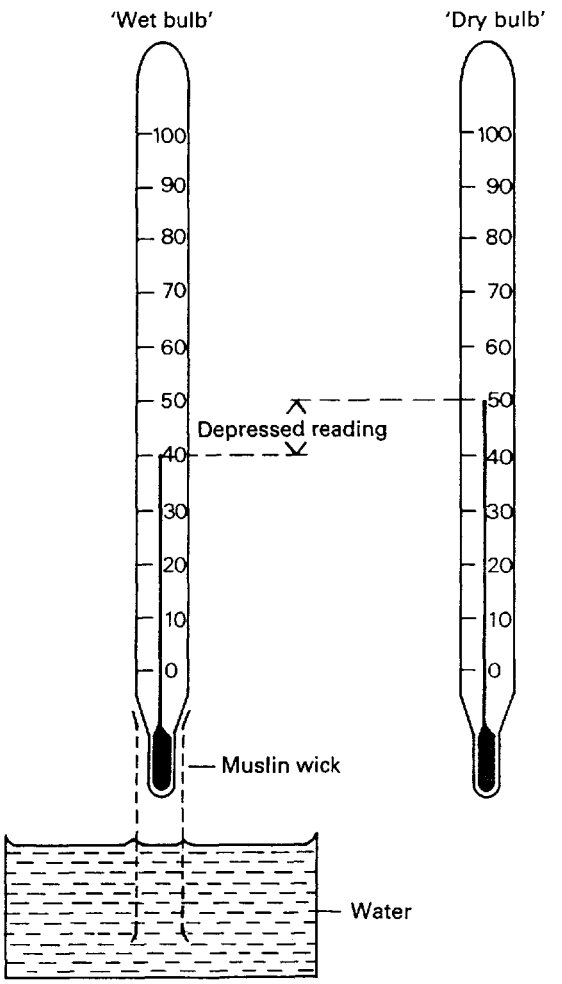


Fig. 26.1 Wet- and dry-bulb temperatures.

bulb. In a similar way, the temperature of a wet solid is kept low while free water remains, but rises towards the temperature of the drying air as drying proceeds. Fortunately, many materials can withstand higher temperatures when in the dry state.

The two temperatures will be the same only when the relative humidity is 100%, as under these conditions there will be no net evaporation of water from the sleeve. In fact, the relationship between dry-bulb temperature and the depression of the wet-bulb temperature as a result of evaporation is so precise that these readings can be used to calculate the percentage relative humidity of the air with some accuracy.

Moisture content of air

The moisture content of air, which is the moisture content expressed as kg of water per kg of 'bone-dry'

air, should be carefully distinguished from the relative humidity. Moisture content is not altered by change of temperature alone, only if further moisture is taken up by the air.

Relationship between equilibrium moisture content and relative humidity

The equilibrium moisture content of a solid exposed to moist air varies with the relative humidity, as shown in some typical plots (Fig. 26.2). Ordinary atmospheric conditions are of the order of 20°C and 70–75% relative humidity, so that if exposed to the atmosphere a mineral such as kaolin will contain about 1% moisture, whereas a starch-based product may have as much as 30% or more. Materials exposed to humid conditions will regain moisture, and so there is no advantage in drying to a moisture content lower than that which the material will have under the conditions of use.

Loss of water from wet solids

As explained above, unbound water is easily lost by evaporation until the equilibrium moisture content of the solid is reached. This is shown in Figure 26.3. Once the solid reaches its equilibrium moisture content, extending the time of drying will not change the moisture content as an equilibrium situation has been reached. The only way to reduce the moisture content of the solid shown in Figure 26.3 is to reduce the relative humidity of the ambient air. This can be done mechanically on a large scale with an air-conditioning system. On a small scale, desiccators are used. Silica gel (a common laboratory desiccant) does not directly take water from a solid, instead, it acts by removing the water from the air, thereby reducing its relative humidity to around 5–10%. This in turn moves the drying curve in Figure 26.3 to the left, thus reducing the moisture content of the solids in the desiccator. Phosphorous pentoxide works in an identical manner but it has an even greater affinity for the water in the storage air.

It is worth re-emphasizing at this point that moisture may be regained very quickly from the atmosphere if a 'dry' solid is exposed to ambient air. For this reason it is unnecessary to 'overdry' a product. If a low residual moisture content is necessary because of hydrolytic instability in the material, the dried product must be efficiently sealed from the ingress of moisture. It also worthy of note that some materials (tablet granules are a classic example) have superior compaction properties with a small amount (1–2%) of residual moisture.

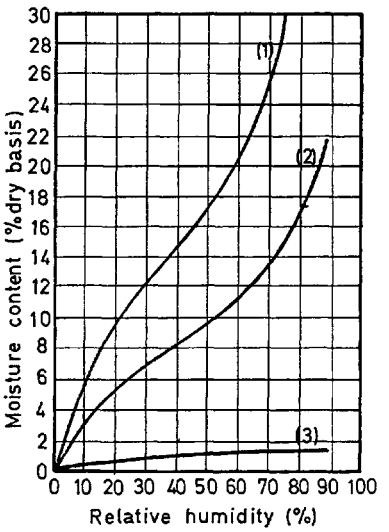


Fig. 26.2 Typical equilibrium moisture contents at 20°C. (1) Starch-based materials; (2) textiles and fibrous materials; (3) inorganic substances such as kaolin.

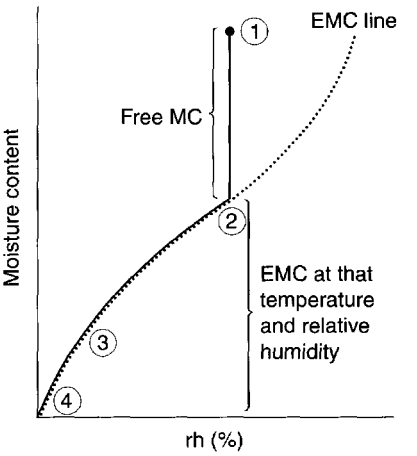


Fig. 26.3 Loss of water from a drying solid. The wet solid prior to drying is at condition (1). It can lose water by evaporation to position (2), its equilibrium moisture content at that RH. The only way the solid can lose more water is to reduce the RH of the ambient or storage atmosphere, to (3) with silica gel or to (4) with phosphorus pentoxide.

TYPES OF DRYING METHOD

When considering how to dry a material, the following points should be considered:

- Heat sensitivity of the material being dried

- Physical characteristics of the material
- The necessity for asepsis
- Nature of the liquid to be removed
- The scale of the operation
- Available sources of heat (steam, electrical).

The general principles for efficient drying can be summarized as follows:

- Large surface area for heat transfer;
- Efficient heat transfer per unit area (to supply sufficient latent heat of vaporization or heat of sublimation in the case of freeze-drying);
- Efficient mass transfer of evaporated water through any surrounding boundary layers, i.e. sufficient turbulence to minimize boundary layer thickness;
- Efficient vapour removal, i.e. low relative humidity air at adequate velocity.

It is convenient to categorize pharmaceutical driers according to the heat transfer method they use, i.e. convective, conductive or radiant.

CONVECTIVE DRYING OF WET SOLIDS

Fixed (or static) bed convective drying

The factors affecting drying in this manner can be illustrated by reference to the construction and use of a simple form of drier – the tray (shelf or compartment) drier.

Tray drier

An efficient type of tray drier is the directed circulation form shown in Figure 26.4. Air flows in the direction of the arrows over each shelf in turn. The wet material is spread on shallow trays resting on the shelves. Electrical elements or steam-heated pipes are positioned as shown, so that the air is periodically reheated after it has cooled by passage over the wet material on one shelf before it passes over the material on the next.

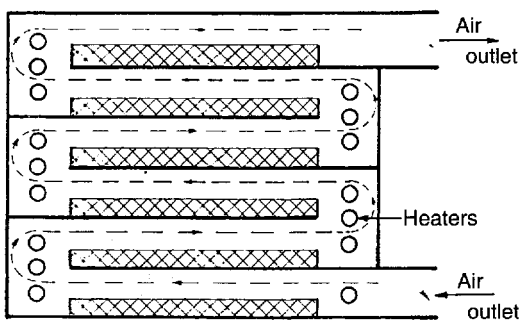


Fig. 26.4 Directed-circulation tray drier.

The required latent heat of evaporation is transferred convectively from the air and the rate of heat transfer may be written as:

$$\text{Rate of heat transfer, } (dH/dt) = h_c A\ \Delta T$$

where h_c is a heat transfer coefficient (Chapter 38) for convective heat transfer. The value of h_c is commonly around only 10–20 W m^{-2} K^{-1}. Heat transfer from air is therefore relatively inefficient and so convective drying is slow and wet materials can take up to 24 hours to dry.

There is another important factor controlling the rate of drying: the water vapour must pass through the boundary layers present at the surface into the turbulent airstream. For this to occur the relative humidity of the air must be kept well below the saturation level and the boundary layers small. These conditions are achieved by having a brisk turbulent air flow over the surface and by the periodic reheating of the air as the temperature falls, so that it can pick up further moisture.

Rate of drying in fixed beds

The rate at which drying occurs has been found to show certain phases (Fig. 26.5) in which the change in moisture content is plotted against time. From A to B the relationship is linear, which is known as the ***constant-rate period***, whereas from B to C the rate of loss of moisture decreases and is known as the ***falling-rate period***. The end of the constant rate period, B, is referred to as the ***critical moisture content***.

The ***first falling-rate period*** has a linear relationship, that is, the decrease in drying rate is uniform, whereas in the ***second falling-rate period*** there is a continuous decrease in the rate of drying until the equilibrium moisture content is reached. Each of these periods will be considered in more detail.

Constant-rate period For given conditions of temperature and humidity, most substances dry at a similar rate in the constant-rate period. It is found that the evaporation rate from the drying bed is similar to that of the solvent alone from a free liquid surface under the same conditions, indicating that the evaporation takes place from the wet surface of the solid, and that the surface remains wet in this period as a result of the liquid being replaced from below as fast as it is vaporized.

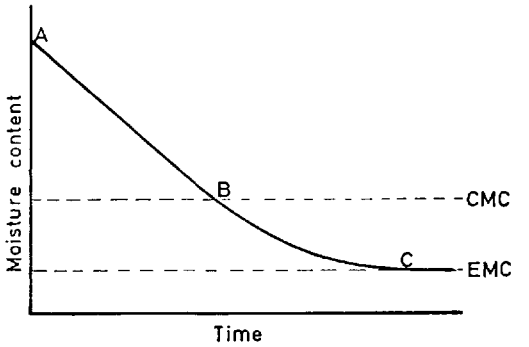


Fig. 26.5 Drying curve. CMC, critical moisture content, EMC, equilibrium moisture content.

Controlling factors in this period are the rate at which heat can be transferred and the rate of removal of the vapour, as explained earlier.

First falling-rate period As moisture is removed from the surface, a point will be reached when the rate of vaporization is insufficient to saturate the air in contact with the surface. Under these conditions, the rate of drying will be limited by the rate of capillary transfer of the liquid to the surface of the wet bed, and as this becomes increasingly difficult as the bed dries, the solvent level decreases and thus has further to travel to the point of evaporation. Consequently, the rate of drying decreases continuously.

Moisture movement may cause the 'migration' of soluble drugs or excipients. This is discussed later in this chapter.

Eventually, movement of solvent to the surface can no longer occur, for example when the water is in the pendular state (Fig. 25.2) and drying at the surface will end. As the drying rate decreases, less heat is used as latent heat of vaporization, so that the heat input should be reduced.

Second falling-rate period Any moisture that remains within the drying bed at the end of the first falling-rate period is unable to move, so that drying cannot take place on the surface. Hence, the plane of vaporization retreats from the surface into the body of the solid, and the drying rate depends on the movement of the ***vapour*** through the pores of the bed to the surface, in general by molecular diffusion.

Minimal atmospheric humidity above the solid will assist in maintaining the maximum vapour pressure gradient. In addition, the thermal conductivity of the solid decreases as it becomes dry; if the solid is thermostable it is safe to allow temperature gradients to increase to maintain the rate of heat transfer, but if the material is thermolabile the heating must be decreased.

In the operation of a tray drier it is usual to remove the dry material on the trays near the air inlet and replace them with the trays with partially dry material from further away. Trays with fresh wet material are placed on the empty shelves. In this way the outgoing (wetter) air contacts the wettest material.

Dynamic convective driers

Fluidized-bed drier

An excellent method of obtaining good contact between the warm drying air and wet particles is found in the ***fluidized-bed drier***. The general principles of ***fluidization*** will be summarized before discussing its application to drying.

Consider the situation in which particulate matter is contained in a vessel, the base of which is perforated, enabling a fluid to pass through the bed of solids from below. The fluid can be liquid or gas, but for the purposes of this description air will be assumed, as it is directly relevant to the drying process.

If the air velocity through the bed is increased gradually and the pressure drop through the bed is measured, a graph of the operation shows several distinct regions, as indicated in Figure 26.6. At first, when the air velocity is low, A, flow takes place between the particles without causing disturbance, but as the velocity is increased a point, B, is reached, when the pressure drop has attained a value where the frictional drag on the particle is equal to the force of gravity on that particle. Rearrangement of the particles occurs to offer least resistance, C, and eventually they are suspended in the air and can move; pressure drop through the bed decreases slightly because of the greater porosity, D. Further increase in the air velocity causes the particles to separate and move freely and the bed is ***fully fluidized***. Any additional increase in velocity separates the particles

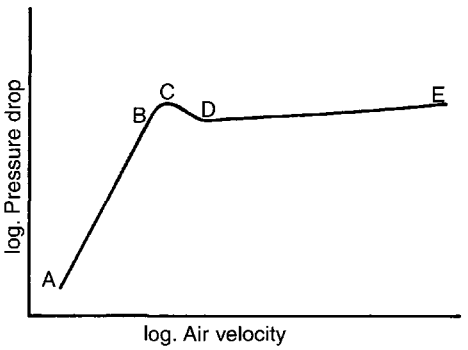


Fig. 26.6 Effect of air velocity on pressure drop through a fluidized bed.

further, that is, the bed expands without appreciable change in the pressure drop, until E, when the air velocity is sufficient to entrain the solid particles and transport them out of the top of the bed.

In the region D–E fluidization is irregular, much of the air flowing through in bubbles, the term ***boiling bed*** being commonly used to describe it. The important fact is that it produces conditions of great turbulence, the particles mixing with good contact between them and the air. Hence if hot air is used the turbulent conditions lead to high heat and mass transfer rates; the fluidized-bed technique therefore offers a means of rapid drying. The arrangement of such a drier is shown in Figure 26.7. Sizes are available with capacities from 1 kg in the laboratory to 200–500 kg in production.

Advantages of fluidized-bed drying

1. Efficient heat and mass transfer give high drying rates, so that drying times are shorter than with static-bed convection driers. A batch of tablet granules, for example, can be dried in 20–30 minutes, whereas a tray drier would require many hours. Apart from obvious economic advantages, the heat challenge to thermolabile materials is minimized.
2. The fluidized state of the bed ensures that drying occurs from the surface of all the individual particles and not just from the surface of the bed. Hence, most of the drying will be at constant rate and the failing-rate period (when the danger of overheating is greatest) is very short.
3. The temperature of a fluidized bed is uniform throughout and can be controlled precisely.
4. The turbulence in a fluidized bed causes some attrition to the surface of the granule. This produces a more spherical free-flowing product.
5. The free movement of individual particles eliminates the risk of soluble materials migrating, as may occur in static beds (see later).
6. The containers can be mobile, making handling and movement around the production area simple and so reducing labour costs.
7. Short drying times mean that the unit has a high output from a small floor space.

Disadvantages of fluidized-bed drying

1. The turbulence of the fluidized state may cause excessive attrition of some materials, with damage to some granules and the production of too much dust.
2. Fine particles may become entrained in the fluidizing air and must be collected by bag filters, with care to avoid segregation and loss of fines.
3. The vigorous movement of particles in hot dry air can lead to the generation of static electricity charges, and suitable precautions must be taken. A mixture of air with a fine dust of organic materials such as starch and lactose can explode

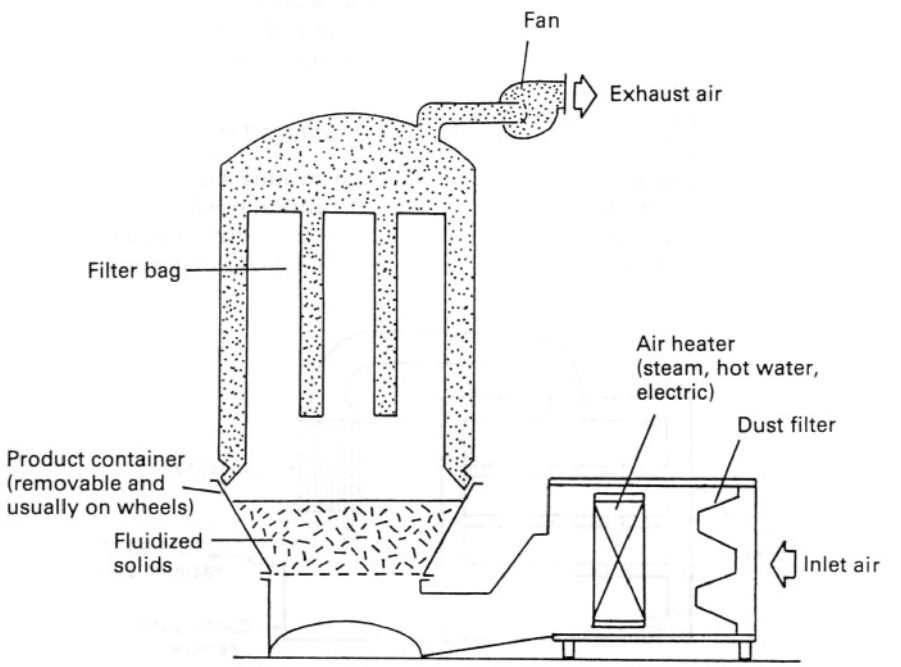


Fig. 26.7 Fluidized-bed drier.

violently if ignited by sparking caused by static charges. The danger is increased if the fluidized material contains a volatile solvent such as isopropanol. Adequate electrical earthing is essential.

CONDUCTIVE DRYING OF WET SOLIDS

In this process the wet solid is in thermal contact with a hot surface and the bulk of heat transfer occurs by conduction.

Vacuum oven

This equipment is a good example of a conduction drier though it is not used so extensively as it was formerly. The vacuum oven (Fig. 26.8) consists of a jacketed vessel sufficiently stout in construction to withstand vacuum within the oven and steam pressure in the jacket. In addition, the supports for the shelves form part of the jacket, giving a larger area for conduction heat transfer. The oven can be closed by a door that can be locked to give an airtight seal. The oven is connected through a condenser and liquid receiver to a vacuum pump, although if the liquid to be removed is water and the pump is of the ejector type that can handle water vapour, the pump can be connected directly to the oven.

Operating pressure can be as low as 0.03–0.06 bar, at which pressures water boils at 25–35°C. Some ovens may be large (for example about 1.5 m cube and with 20 shelves), but vacuum ovens are rarely used nowadays for production, although they are frequently found in development laboratories, where they are commonly used for the drying of small development samples, particularly when the heat stability of the drug or formulation is uncertain.

The main advantage of a vacuum oven is that drying takes place at a low temperature, and as there is little air present there is minimum risk of oxidation. The temperature of the drying solid will rise to the steam or water temperature at the end of the drying, but this is not usually harmful.

Vacuum tumbling drier

Vacuum tumbling drying has found application in the pharmaceutical industry. One design of tumbler drier resembles a large Y-cone mixer (discussed in Chapter 13). The vessel is steam jacketed and is connected to a vacuum. It can be used for drying tablet granules, which tumble over the heated surface as the vessel slowly revolves. Heat transfer rates in this equipment are much higher than can be attained in a conventional vacuum oven, where the material is static.

RADIATION DRYING OF WET SOLIDS

Radiant heat transmission

Heat transmission by radiation differs from heat transfer by conduction or convection in that no transfer medium (solid, liquid or gaseous) need be present. Heat energy in the form of radiation can cross empty space or travel through the atmosphere virtually without loss. If it falls on a body capable of absorbing it then it appears as heat, although a proportion may be reflected or transmitted.

Use of infrared radiation

Infrared heating has been used in the past to dry pharmaceutical products such as wet granules, but it suffers from the disadvantage that it is absorbed very quickly and does not penetrate far into the wet mass.

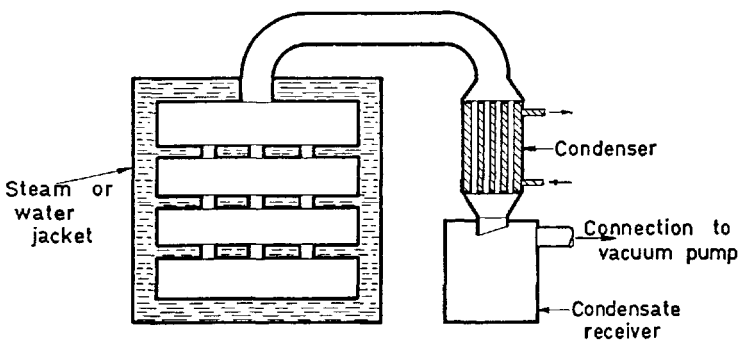


Fig. 26.8 Vacuum oven (schematic).

The surface layers dry quickly and the absorption of further energy then raises the temperature of the dry material to a high value, which is often detrimental to the product. For this reason infrared radiation is now seldom used as a heat source in pharmaceutical manufacture.

The use of microwave radiation

Although energy in the infrared region is more easily generated there are other, longer, wavelengths that can generate heat when the radiation is absorbed by a wet solid. Microwave radiation in the wavelength range 10 mm to 1 m penetrates much better than IR radiation. Microwave driers are finding some application in the pharmaceutical industry.

Table 26.1 Microwave energy loss factors for some pharmaceutical solvents and excipients

Material	Loss factor
Methanol	13.6
Ethanol	8.6
Water	6.1
Isopropanol	2.9
Acetone	1.25
Maize starch	0.41
Magnesium carbonate	0.08
Lactose	0.02

Generation and action of microwaves

Microwaves are produced by an electronic device known as a magnetron. Microwave energy can be reflected down a rectangular duct (termed a waveguide) or simply beamed through a transparent polypropylene window into the drying chamber. To avoid interference with radio and television it is permitted to operate only at certain frequencies, which are normally 960 and 2450 MHz.

The penetration of microwaves into the wet product is so good that heat is generated uniformly within the solid.

When microwaves fall on substances of suitable electronic structure (small polar molecules, such as water), the electrons in the molecule attempt to resonate in sympathy with the radiation and the resulting molecular 'friction' results in the generation of heat. Dry solids do not resonate as well as water, so further heating may be avoided once the water is removed. This is indicated clearly by the loss factors listed in Table 26.1. The loss factor is a measure of the ratio of the microwave energy absorbed by individual molecules; the higher the number the greater the absorption of microwave energy. Table 26.1 lists these values for some common solvents and excipients. Clearly, the absorption of the microwave energy is far greater for small polar molecules than for larger and less polar molecules.

A microwave drier for granulates

Figure 26.9 is a sketch of a microwave drier used for drying granulates. It is designed to operate under a slight vacuum. This in itself is not essential for the use of microwaves, but the air flow allows the continuous removal of evaporated solvent. The radiation is generated by multiple magnetrons, each producing 0.75 kW at 2450 MHz. The radiation passes through the polypropylene window into the drying chamber, where it is absorbed by the liquid in the wet granules contained on a tray. The heat generated in the mass drives off the moisture and the evolved vapour is drawn away in the air flow as it is formed. When drying is nearly complete the radiation field intensity will rise, as the dry solids do not absorb as readily as water. This rise is detected and the magnetrons are progressively turned off automatically, to give an accurate control of the final moisture content and minimize the danger of overheating.

Advantages of microwave drying The following advantages are claimed for microwave drying:

1. It provides rapid drying at fairly low temperatures.
2. The thermal efficiency is high, as the drier casing and the air remain cool. Most of the microwave energy is absorbed by the liquid in the wet material.
3. The bed is stationary, avoiding the problems of dust and attrition.
4. Solute migration is reduced as there is uniform heating of the wet mass.
5. Equipment is highly efficient and refined. All the requirements of product and operator safety have been incorporated into machines without detracting from GMP considerations.
6. Granulation end-point is possible by measuring the residual microwave energy (as this rises sharply when there is little solvent left to evaporate).

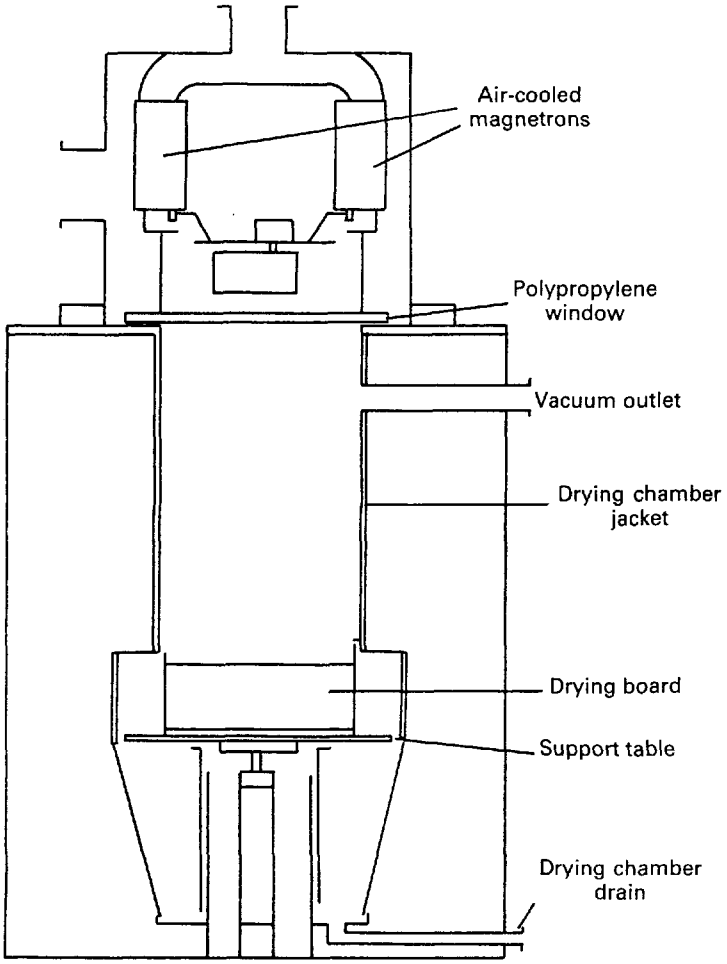


Fig. 26.9 Microwave drier (courtesy of T K Fielder).

Disadvantages of microwave drying

1. The batch size of commercial production microwave driers is smaller than those available for fluidized-bed driers.
2. Care must be taken to shield operators from the microwave radiation, which can cause damage to organs such as the eyes and testes. This is ensured by 'failsafe' devices preventing the generation of microwaves until the drying chamber is sealed.

DRIERS FOR DILUTE SOLUTIONS AND SUSPENSIONS

The objective of these driers is to spread the liquid to a large surface area for heat and mass transfer and to provide an effective means of collecting the dry solid. Two main types are used, the first spreading the liquid to a thin film on to a drum and the second dispersing the liquid to a spray of small droplets.

Drum drier

Shown in section in Figure 26.10, the drum drier consists of a drum 0.75–1.5 m in diameter and 2–4 m in length, heated internally, usually by steam, and rotated on its longitudinal axis. The liquid is applied to the surface of the drum and spread to a film; this may be done in various ways, but the simplest method is that shown in the diagram, where the drum dips into a **feed pan**. Drying rate is controlled by manipulating the speed of rotation of the drum and its temperature. The drum can be heated by either steam or warm water. The product is scraped

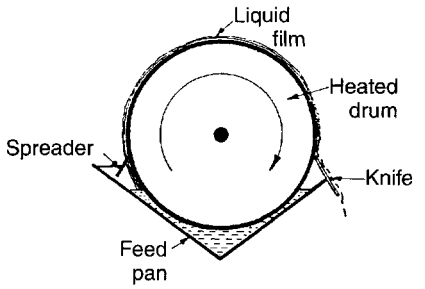


Fig. 26.10 Drum drier.

from the surface of the drum by means of a ***doctor knife***.

Advantages of the drum drier

1. The method gives rapid drying, the thin film spread over a large area resulting in rapid heat and mass transfer.
2. The equipment is compact, occupying much less space than the spray-drier, for example.
3. Heating time is short, being only a few seconds.
4. The drum can be enclosed in a vacuum jacket, enabling the temperature of drying to be reduced.
5. The product is obtained in flake form, which is convenient for many purposes.

The only ***disadvantage*** is that operating conditions are critical and it is necessary to impose careful control on feed rate, film thickness, speed of drum rotation and drum temperature.

The drum drier can handle a variety of materials, either as solutions or as suspensions; substances that are dried by this method include starch products, ferrous salts and suspensions of kaolin and zinc oxide.

Spray drier

The spray drier provides a large surface area for heat and mass transfer by atomizing the liquid to small droplets. These are sprayed into a stream of hot air, so that each droplet dries to an individual solid particle.

There are many forms of spray drier and Figure 26.11 shows a typical design, in which the drying chamber resembles a cyclone. This ensures good air circulation, facilitates heat and mass transfer and encourages the separation of dried particles from the moving air by the centrifugal action.

The character of the particles is controlled by the droplet size, and so the type of atomizer is important. Jet atomizers are easily blocked by rapid evaporation and deposition of solid on the nozzle, and the droplet size is likely to vary. This is not the case with

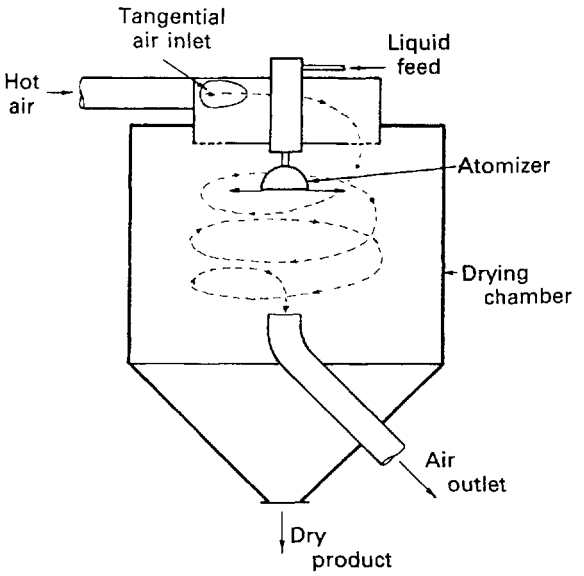


Fig. 26.11 Spray drier.

rotary types of atomizer, one form of which is shown in Figure 26.12. Liquid is fed on to the disc, which is rotated at high speed (up to 20 000 rpm). A film is formed and spreads from the small disc to a larger, inverted hemispherical bowl, becoming thinner and eventually being dispersed from the edge in a fine, uniform spray. In addition, the rotary atomizer has the advantage of being equally effective with either solutions or suspensions of solids, and it can operate efficiently at various feed rates.

The air enters the chamber tangentially and rotates the drying droplets around the chamber to increase their residence time and therefore time for drying. For pharmaceutical purposes it is usual to filter the air and to heat it indirectly by means of a heat exchanger. Dust carried over in the air outlet stream may be recovered by a cyclone separator or filter bag.

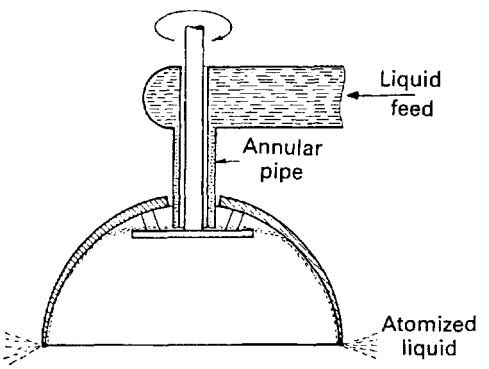


Fig. 26.12 Rotary atomizer.

Spray-dried products are easily recognizable, being uniform in appearance. The particles have a characteristic shape, in the form of hollow spheres sometimes with a small hole. This arises from the drying process, as the droplet enters the hot air stream and dries on the outside to form an outer crust with liquid still in the centre. This liquid then vaporizes, and the internal vapour escapes by blowing a hole in the sphere. Figure 26.13 shows the mechanism of formation of the spherical product.

Advantages of the spray drying process

1. There are millions of small droplets which give a large surface area for heat and mass transfer, so that evaporation is very rapid. The actual drying time of a droplet is only a fraction of a second, and the overall time in the drier only a few seconds.
2. Because evaporation is very rapid, the droplets do not attain a high temperature. Most of the heat is used as latent heat of vaporization and so the temperature of the particles is kept low by evaporative cooling.
3. The characteristic particle form gives the product a high bulk density and, in turn, rapid dissolution (large surface area).
4. Provided that a suitable atomizer is used the resulting powder will have a uniform and controllable particle size.
5. The product is free flowing, with almost spherical particles, and is especially convenient for tablet manufacture as it has excellent flow and compaction properties.
6. Labour costs are low, the process yielding a dry, free-flowing powder from a dilute solution in a single operation with no handling.

Disadvantages of the spray drying process

1. The equipment is very bulky, and with the ancillary equipment is expensive. In a large installation the drying chamber alone may be as much as 15 m in height and 6 m in diameter.
2. The overall thermal efficiency is rather low, as the air must still be hot enough when it leaves the drier to avoid condensation of moisture. Also, large volumes of heated air pass through the chamber without contacting a particle, thus not contributing directly to the drying process.

Uses of the spray drying process The spray drier can be used for drying almost any substance, in solution or in suspension. It is most useful for thermolabile materials, particularly if handled continuously and in large quantities; outputs of 2000 kg h^{-1} can be attained, although pharmaceutical plants are usually somewhat smaller.

Examples of both soluble and insoluble substances that are spray dried include citric acid, sodium phosphate gelatin, starch, barium sulphate, calcium phosphate, and some powdered antibiotic formulations for reconstitution into syrup.

Spray drying is also capable of producing spherical particles in the respirable range of 1–7 mm that have been used satisfactorily for the delivery of drugs from dry powder inhalers.

It is possible to operate spray driers aseptically using heated filtered air to dry products such as serum hydrolysate. Also, some spray driers operate in a closed-circuit mode with an inert gas to minimize oxidation of the product. Volatile solvents can be recovered from such systems.

Modern pharmaceutical spray drying has been reviewed by Wendel and Çelik (1997) and the reader is referred to this article if additional information is required.

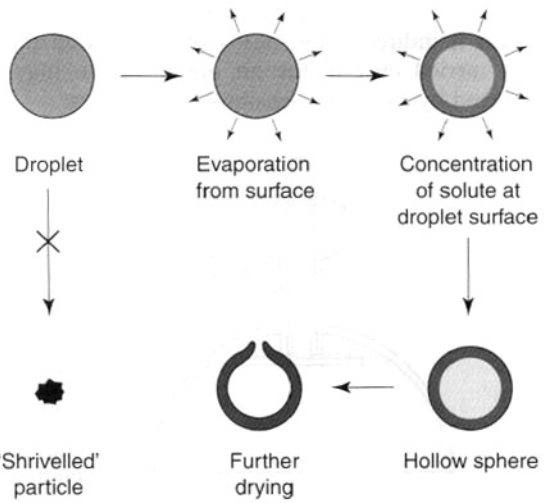


Fig. 26.13 Formation of product in spray drying.

FREEZE DRYING

Freeze drying is a process used to dry extremely heat-sensitive materials. It allows the drying, without excessive damage, of proteins, blood products and even microorganisms, which retain a small but significant viability.

In this process the initial liquid solution or suspension is frozen, the pressure above the frozen state

is reduced and the water removed by sublimation. Thus a liquid-to-vapour transition takes place, as with all the previous driers discussed, but here there are three states of matter involved: liquid to solid, then solid to vapour.

The theory and practice of freeze drying is based, therefore, on an understanding and application of the phase diagram for the water system.

The phase diagram for water

The phase diagram for the water system is shown in Figure 26.14. The diagram consists of three separate areas, each representing a single phase of water, either solid, liquid or vapour. Two phases can coexist along a line under the conditions of temperature and pressure defined by any point on the line. The point O is the one unique point where all three phases can coexist, and is known as the ***triple point***. Its coordinates are a pressure of 610 Pa and a temperature of 0.0075°C.

The lines on the phase diagram represent the interphase equilibrium lines, which show:

1. the boiling point of water as it is lowered by reduction of the external pressure above the water (BO in Fig. 26.14);
2. the variation of the melting point of ice on reduction of the external pressure above it. There is a very slight rise in the melting point (AO);
3. the reduction of the vapour pressure exerted by ice as the temperature is reduced (CO).

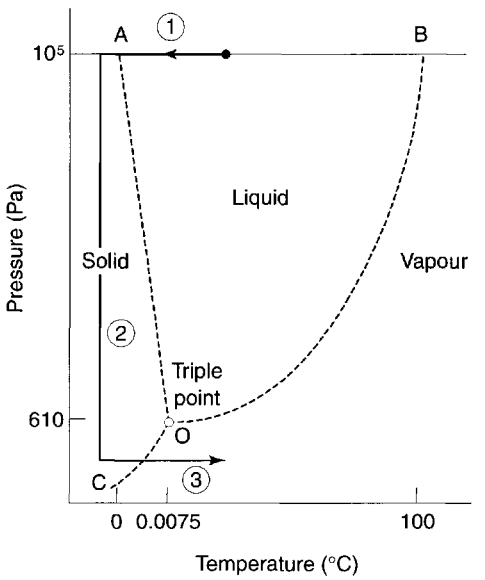


Fig. 26.14 The phase diagram for water (not to scale) with freeze drying process superimposed (see text for explanation).

On heating at constant ***atmospheric*** pressure ice will melt when the temperature rises to 0°C. At this constant temperature and pressure it will then change to water. Continued heating will raise the temperature of the water to 100°C where, if heat addition is continued, the liquid water will be converted into water vapour at 100°C.

If, however, solid ice is maintained at a pressure below the triple point then on heating the ice will sublime and pass directly to water vapour without passing through the liquid phase. This sublimation, and therefore drying, can occur at a temperature below 0°C. This will only happen if the pressure is prevented from rising above the triple point pressure and, to ensure that this is the case, the vapour evolved must be removed as fast as it is formed. It may be thought that as the process takes place at a low temperature the heat required to sublime the ice will be small. In fact, the latent heat of sublimation of ice at 2900 kJ kg^{-1} is appreciably greater than the latent heat of evaporation of water at atmospheric pressure, and this heat must be supplied for the process to take place.

Application of the phase diagram of water to freeze drying

The freeze drying of products such as blood plasma, although simple in theory, presents a number of practical problems.

1. The depression of the freezing point caused by the presence of dissolved solutes means that the solution must be cooled to well below the normal freezing temperature for pure water, and it is usual to work in the range −10 to −30°C. In part this is because it is obviously not pure water that is being dried, and thus the presence of dissolved solutes will shift the pure-water phase diagram.
2. Sublimation can only occur at the frozen surface and is a slow process (approximately 1 mm thickness of ice per hour). For all but very small volumes the surface area must therefore be increased and the liquid thickness prior to freezing be reduced in order to reduce the thickness of ice to be sublimated.
3. At low pressures large volumes of water vapour are produced which must be rapidly removed to prevent the pressure rising above the triple point pressure.

4. The dry material often needs to be sterile, and it must also be prevented from regaining moisture prior to final packing.

Stages of the freeze drying process

Freezing stage

The liquid material is frozen before the application of vacuum to avoid frothing, and several methods are used to produce a large frozen surface.

Shell freezing This is employed for fairly large volumes such as blood products. The bottles are rotated slowly and almost horizontally in a refrigerated bath. The liquid freezes in a thin shell around the inner circumference of the bottle. Freezing is slow and large ice crystals form, which is a drawback of this method as they may damage blood cells and reduce the viability of microbial cultures.

In vertical spin freezing the bottles are spun individually in a vertical position so that centrifugal force forms a circumferential layer of solution, which is cooled by a blast of cold air. The solution supercools and freezes rapidly, with the formation of small ice crystals.

Centrifugal evaporative freezing This is a similar method, where the solution is spun in small containers within a centrifuge. This prevents foaming when a vacuum is applied. The vacuum causes boiling at room temperature and this removes so much latent heat that the solution cools quickly and snap freezes. About 20% of the water is removed prior to freeze drying and there is no need for refrigeration. Ampoules are usually frozen in this way, a number being spun in a horizontal angled position in a special centrifuge head so that the liquid is thrown outwards and freezes as a wedge.

Vacuum application stage

The containers and the frozen material must be connected to a vacuum source sufficient to drop the pressure below the triple point and remove the large volumes of low-pressure vapour formed during drying. Again an excess vacuum is normal in practice, to ensure that the product in question is below its triple point.

Commonly a number of bottles or vials are attached to individual outlets of a manifold, which is connected to a vacuum.

Sublimation stage

Heat of sublimation must be supplied. Under these conditions the ice slowly sublimes, leaving a porous solid which still contains about 0.5% moisture after primary drying.

Primary drying Primary drying can reduce the moisture content of a freeze-dried solid to around 0.5%. Further reduction can be effected by secondary drying. During the primary drying, the latent heat of sublimation must be provided and the vapour removed.

Heat transfer Heat transfer is critical: insufficient heat input prolongs the process, which is already slow, and excess heat will cause melting.

Prefrozen bottles – of blood, for example – are placed in individually heated cylinders, or are connected to a manifold when heat can be taken from the atmosphere.

Shelf-frozen materials are heated from the drier shelf, whereas ampoules may be left on the centrifuge head or may be placed on a manifold, but in either case heat from the atmosphere is sufficient.

In all cases the heat transfer must be controlled, as only about 5 W m^{-2} K^{-1} is needed and overheating will lead to melting. It is important to appreciate here that although a significant amount of heat is required there should be no significant increase in temperature – the added heat should be sufficient to provide the latent heat of sublimation only and little sensible heat.

Vapour removal The vapour formed must be continually removed to avoid a pressure rise that would stop sublimation. To reduce the pressure sufficiently it is necessary to use efficient vacuum pumps, usually two-stage rotary pumps on the small scale, and ejector pumps on the large scale. On the small scale, vapour is absorbed by a desiccant such as phosphorus pentoxide, or is cooled in a small condenser with solid carbon dioxide. Mechanically refrigerated condensers are used on the large scale.

For vapour flow to occur the vapour pressure at the condenser must be less than that at the frozen surface, and a low condenser temperature is necessary. On the large scale vapour is commonly removed by pumping, but the pumps must be of large capacity and not affected by moisture. The extent of the necessary pumping capacity will be realized from the fact that, under the pressure conditions used during primary drying, 1 g of ice will form 1000 L of water vapour. Ejector pumps are most satisfactory for this purpose.

Rate of drying The rate of drying in freeze drying is very slow, the ice being removed at a rate of about only 1 mm depth per hour. The drying rate curve illustrated in Figure 26.15 shows a similar shape to a normal drying curve, the drying being at constant rate during most of the time.

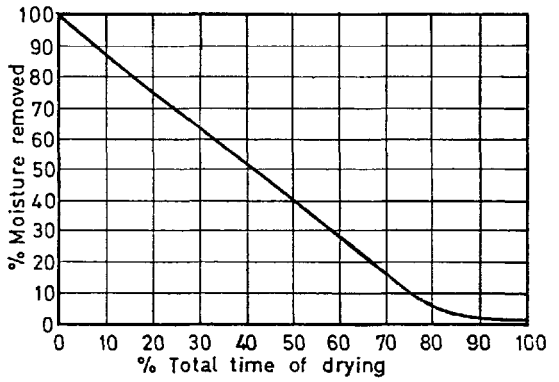


Fig. 26.15 Sublimation drying: rate of drying curve.

Computer control enables the drying cycle to be monitored. There is an optimum vapour pressure for a maximum sublimation rate and the heat input and other variables are adjusted to maintain this value. Continuous freeze drying is possible in modern equipment, where the vacuum chamber is fitted with a belt conveyor and vacuum locks, but despite these advances the overall drying rate is still slow.

Secondary drying

The removal of residual moisture at the end of primary drying is performed by raising the temperature of the solid to as high as 50 or 60°C. A high temperature is permissible for many materials because the small amount of moisture remaining is not sufficient to cause spoilage.

Packaging

Attention must be paid to packaging freeze-dried products to ensure protection from moisture. Containers should be closed without contacting the atmosphere, if possible, and ampoules, for example, are sealed on the manifold while still under vacuum. Otherwise, the closing must be carried out under controlled atmospheric conditions.

Freeze drying in practice

Advantages

As a result of the character of the process, freeze drying has certain special advantages:

1. Drying takes place at very low temperatures, so that enzyme action is inhibited and chemical decomposition, particularly hydrolysis, is minimized.
2. The solution is frozen such that the final dry product is a network of solid occupying the same volume as the original solution. Thus, the product is light and porous.
3. The porous form of the product gives ready solubility.
4. There is no concentration of the solution prior to drying. Hence, salts do not concentrate and denature proteins, as occurs with other drying methods.
5. As the process takes place under high vacuum there is little contact with air, and oxidation is minimized.

Disadvantages

There are two main disadvantages of freeze drying:

1. The porosity, ready solubility and complete dryness yield a very hygroscopic product. Unless products are dried in their final container and sealed in situ, packaging requires special conditions.
2. The process is very slow and uses complicated plant, which is very expensive. It is not a general method of drying, therefore, but is limited to certain types of valuable products which, because of their heat sensitivity, cannot be dried by any other means.

Uses of freeze drying

The method is used for products that cannot be dried by any other heat method. These include biological products, for example some antibiotics, blood products, vaccines (such as BCG, yellow fever, smallpox), enzyme preparations (such as hyaluronidase) and microbiological cultures. The latter enables specific microbiological species and strains to be stored for long periods with a viability of about 10% on reconstitution.

SOLUTE MIGRATION DURING DRYING

Solute migration is the phenomenon that can occur during drying which results from the movement of a solution within a wet system. The solvent moves towards the surface of a solid (from where it evaporates), taking any dissolved solute with it. Many drugs and binding agents are soluble in granulating fluid, and during the convective drying of granulates these solutes can move towards the surface of the drying bed

or granule and be deposited there when the solvent evaporates. Solute migration during drying can lead to localized variability in the concentration of soluble drugs and excipients within the dried product.

Migration associated with drying granules can be of two types, intergranular (between granules) and intragranular (within individual granules).

Intergranular migration

Intergranular migration, where the solutes move from granule to granule, may result in gross maldistribution of the active drug. It can occur during the drying of static beds of granules (e.g. tray drying), as the solvent and accompanying solute(s) move from granule to granule towards the top surface of the bed where evaporation takes place. When the granules are compressed the tablets may have a deficiency or an excess of drug. For example, experiments found that only 12% of tablets made from a tray-dried warfarin granulate were within the USP limits for drug content.

Intragranular migration

Drying methods based on fluidization and vacuum tumbling keep the granules separate during drying and so prevent the intergranular migration that may occur in fixed beds. However, intragranular migration, where the solutes move towards the periphery of each granule, may take place.

Consequences of solute migration

Solute migration of either type can result in a number of problems and occasional benefits.

Loss of active drug

The periphery of each granule may become enriched, with the interior suffering a depletion. This will be of no consequence unless the enriched outer layer is abraded and lost, as may happen during fluidized-bed drying, when the fine drug-rich dust can be eluted in the air and carried to the filter bag or lost. The granules suffer a net loss of drug and, as a result, will be below specification with respect to quantity of active ingredient.

Mottling of coloured tablets

Coloured tablets can be made by adding soluble colour during wet granulation. Intragranular migration of the colour may give rise to dry granules with a highly coloured outer zone and a colourless interior (Fig. 26.16). During compaction granules are fractured and the colourless interior is exposed. The eye then sees the coloured fragments against a colourless background and the tablets appear mottled.

Migration may be reduced by using the insoluble aluminium 'lake' of the colouring material (in which the soluble dye is adsorbed strongly on to insoluble alumina particles) in preference to the soluble dye itself. This is not the complete answer, as factors such as an unfavourable pH can allow dyes to detach from lakes and then migrate. They suggest that small granules, which do not fracture so readily, are preferable to larger ones if mottling is troublesome.

Migration of soluble binders

Intragranular migration may deposit a soluble binder at the periphery of the granules and so confer

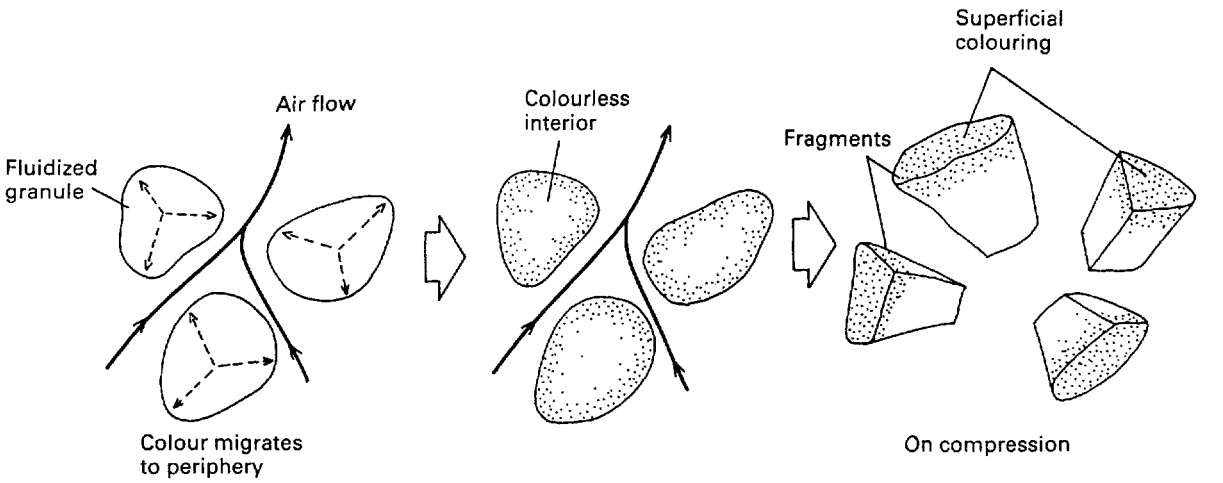


Fig. 26.16 Diagram of mottling caused by intragranular migration.

a 'hoop stress' resistance, making the granules harder and more resistant to abrasion. This migration can aid the bonding process during tablet compaction as a result of binder–binder (rather than drug–drug or drug–excipient) contact, and is therefore sometimes beneficial.

Many other factors, such as granulate formulation, drying method and moisture content, have been shown to affect solute migration.

The influence of formulation factors on solute migration

Nature of substrate

The principles governing solute migration are similar to those of thin-layer chromatography. Thus, if the granule substrate has an affinity for the solute then migration will be impeded. Luckily, many of the common tablet excipients seem to possess this affinity.

It is likely, therefore, that the presence of absorbent materials, such as starch and microcrystalline cellulose, will minimize tablet solute migration.

The use of water-insoluble aluminium lakes (pigments) rather than water-soluble dyes reduces mottling. This effect has also been seen with film-coat colours.

Viscosity of granulating fluid

The popular granulating fluids are solutions of polymers whose viscosity is appreciably greater than that of water alone. This viscosity impedes the movement of moisture by increasing the fluid friction. Increasing the concentration and therefore the viscosity of PVP solution has been shown to slow the migration of drugs in fixed beds of wet granules. Solutions of methylcellulose with comparable viscosities gave similar migration rates, showing that the effect is due to viscosity alone and not to any specific action of either of the binders.

The influence of process factors on solute migration

Drying method

Intergranular migration in fixed beds of granules will occur whenever a particular method of drying creates a temperature gradient. This results in greater evaporation from the hotter zones.

In slow convective drying (e.g. during static tray drying) the maximum concentration of migrated solute will normally occur in the surface of the drying bed, as the process of drying is slow enough to maintain a capillary flow of solvent/solute to the surface over a long period of time.

Drying by microwave radiation results in the uniform heating that is a characteristic of this technique, which in turn minimizes solute migration.

Drying methods that keep the granules in motion will abolish the problem of intergranular migration, but intragranular migration can still occur. This is marked in fluidized granules. Vacuum tumbling methods, on the other hand, greatly reduce migration.

Initial moisture content

The initial moisture content of the granulate will also influence the extent of migration. The greater the moisture content, the greater will be the moisture movement before the pendular state is reached, at which migration cannot continue as there is no longer a continuous layer of mobile liquid water within the wet solid (see Fig. 25.2).

Some practical means of minimizing solute migration

It may be useful to list the measures that can be taken to minimize migration.

1. Use the minimum quantity of granulating fluid and ensure that it is well distributed. High-speed mixer/granulators give better moisture distribution than earlier equipment, and granulates prepared in this way show less migration.
2. Prepare the smallest granules that will flow easily. These are generally satisfactory if mottling is troublesome.
3. Avoid tray drying unless there is no alternative.
4. If tray drying is unavoidable, the dry granules should be remixed before compression. This will ensure that a random mix of enriched and depleted granules will be fed to the tablet machines. This remixing will be more effective if the granule size is small, as there will be a greater number of granules per die fill.
5. If intragranular migration is likely to be troublesome, consider vacuum or microwave drying as an alternative to fluidized-bed drying.

REFERENCES

Wendel, S. and Çelik, M. (1997) An overview of spray-drying applications. *Pharm. Technol.*, **10**, 124–144

BIBLIOGRAPHY

Broadhead, J. et al (1994) The effect of process and formulation variables on the properties of spray-dried beta-galactosidase. *J. Pharm. Pharmacol.*, **46(6)**, 458–467

Masters, K. (1991) *Spray drying handbook*, 5th edn. Longman, Harlow

Nielson, F. (1983) Spray drying of pharmaceuticals. *Mfg Chem. Aerosol News*, **53(7)**, 38–41

Powell, M. (1983) Developments in freeze drying. *Mfg Chem. Aerosol News*, **53(7)**, 47–50

Travers, D.N. (1983) Problems with solute migration. *Mfg Chem. Aerosol News*, **53(3)**, 67–71

Wan, L.S., Heng, P.W., Chia, C.G. (1991) Preparation of coated particles using a spray drying process with an aqueous system. *Int. J. Pharm.*, 77, 183–191

27
Tablets and compaction

Göran Alderborn

CHAPTER CONTENTS

INTRODUCTION

The oral route is the most common way of administering drugs, and among the oral dosage forms tablets of various different types are the most common. Although a variety of tablets exist, with few exceptions (primarily sugar lozenges) tablets are formed by the compression of a powder held within a confined space. The idea of forming a solid dosage form by powder compression is not new. In 1843 the first patent for a hand-operated device used to form a tablet was granted. The use of tablets as dosage form became of interest to the growing pharmaceutical industry, but within pharmacies the pill (a dosage form for oral administration formed by hand into spherical particles about 4–6 mm in diameter) remained the most popular solid dosage form for a long time.

A tablet consists of one or more drugs (active ingredients) as well as a series of other substances used in the formulation of a complete preparation. In the European Pharmacopoeia (3rd edition, 1997) tablets are defined as 'solid preparations each containing a single dose of one or more active ingredients and obtained by compressing uniform volumes of particles. They are intended for oral administration. Some are swallowed whole, some after being chewed, some are dissolved or dispersed in water before being administered and some are retained in the mouth, where the active ingredient is 'liberated'. Thus, a variety of tablets exists and the type of excipients and also the way in which they are incorporated in the tablet vary between the different types. There are also other dosage forms that can be prepared in a similar way, such as suppositories, but which are administered by other routes.

Tablets are used mainly for systemic drug delivery but also for local drug action. For systemic use the drug must be released from the tablet, i.e. normally dissolved in the fluids of the mouth, stomach or intestine, and thereafter be absorbed into the systemic circulation, by which it reaches its site of action. Alternatively, tablets can be formulated for local delivery of drugs in the mouth or gastrointestinal tract, or can be used to increase temporarily the pH of the stomach.

Tablets are popular for several reasons:

- The oral route represents a convenient and safe way of drug administration.
- Compared to liquid dosage forms tablets have general advantages in terms of the chemical and physical stability of the dosage form.
- The preparation procedure enables accurate dosing of the drug.
- Tablets are convenient to handle and can be prepared in a versatile way with respect to their use and to the delivery of the drug.
- Finally, tablets can be mass produced, with robust and quality-controlled production procedures giving an elegant preparation of consistent quality and, in relative terms, low price.

The main disadvantage of tablets as a dosage form concerns the bioavailability of poorly water-soluble or poorly absorbable drugs. In addition, some drugs may cause local irritant effects or otherwise cause harm to the gastrointestinal mucosa.

QUALITY ATTRIBUTES OF TABLETS

Like all other dosage forms, tablets should fulfil a number of specifications regarding their chemical, physical and biological properties. Quality issues relating to the final product are worth considering early in the development process (and thus early in this chapter) as they give an indication of the goal to be achieved during the development and manufacture of tablets.

Tests and specifications for some of these properties are given in pharmacopoeias. The most important of these are dose content and dose uniformity, the release of the drug in terms of tablet disintegration and drug dissolution, and the microbial quality of the preparation. In addition, the authorities and manufacturers define a set of other specifications. One such important property is the resistance of the tablet towards attrition and fracture.

The quality attributes a tablet must fulfil can be summarized as follows:

1. The tablet should include the correct dose of the drug.
2. The appearance of the tablet should be elegant and its weight, size and appearance should be consistent.
3. The drug should be released from the tablet in a controlled and reproducible way.
4. The tablet should be biocompatible, i.e. not include excipients, contaminants and microorganisms that could cause harm to patients.
5. The tablet should be of sufficient mechanical strength to withstand fracture and erosion during handling.
6. The tablet should be chemically, physically and microbiologically stable during the lifetime of the product.

7. The tablet should be formulated into a product acceptable by the patient.
8. The tablet should be packed in a safe manner.

TABLET MANUFACTURING

Stages in tablet formation

Tablets are prepared by forcing particles into close proximity to each other by powder compression, which enables the particles to cohere into a porous, solid specimen of defined geometry. The compression takes place in a die by the action of two punches, the lower and the upper, by which the compressive force is applied. Powder compression is defined as the reduction in volume of a powder owing to the application of a force. Because of the increased proximity of particle surfaces accomplished during compression, bonds are formed between particles which provides coherency to the powder, i.e. a compact is formed. Compaction is defined as the formation of a porous specimen of defined geometry by powder compression.

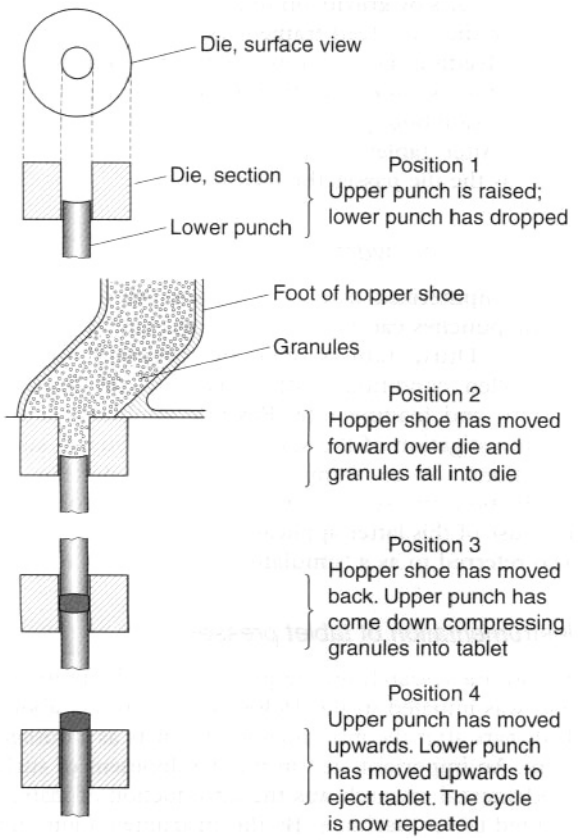


Fig. 27.1 The sequence of events involved in the formation of tablets.

The process of tabletting can be divided into three stages (sometimes known as the ***compaction cycle***) (Fig. 27.1).

Die filling

This is normally accomplished by gravitational flow of the powder from a hopper via the die table into the die (although presses based on centrifugal die filling are also used). The die is closed at its lower end by the lower punch.

Tablet formation

The upper punch descends and enters the die and the powder is compressed until a tablet is formed. During the compression phase, the lower punch can be stationary or can move upwards in the die. After maximum applied force is reached, the upper punch leaves the powder, i.e. the decompression phase.

Tablet ejection

During this phase the lower punch rises until its tip reaches the level of the top of the die. The tablet is subsequently removed from the die and die table by a pushing device.

Tablet presses

There are two types of press in common use during tablet production: the single-punch press and the rotary press. In addition, in research and development work hydraulic presses are used as advanced equipment for the evaluation of the tabletting properties of powders and the prediction of scale-up on the properties of the formed tablets (scale-up refers to the change to a larger apparatus for performing a certain operation on a larger scale).

Single-punch press (eccentric press)

A single-punch press possesses one die and one pair of punches (Fig. 27.2). The powder is held in a hopper which is connected to a hopper shoe located at the die table. The hopper shoe moves to and fro over the die, by either a rotational or a translational movement. When the hopper shoe is located over the die, the powder is fed into the die by gravity. The amount of powder filled into the die is controlled by the position of the lower punch. When the hopper shoe is located beside the die, the upper punch descends and the powder is compressed. The lower punch is stationary during compression and the pressure is thus applied by the upper punch and

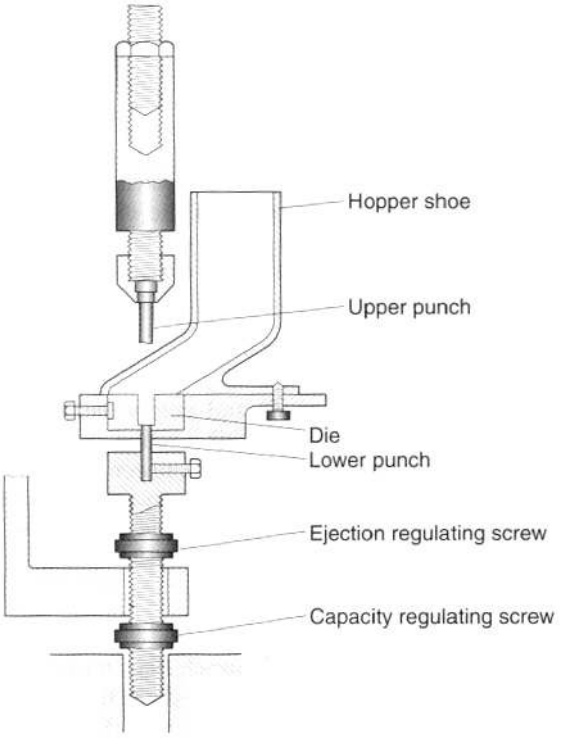


Fig. 27.2 A single-punch tablet press.

controlled by the upper punch displacement. After ejection the tablet is pushed away by the hopper shoe as it moves back to the die for the next tablet.

The output of tablets from a single-punch press is about 200 tablets per minute. A single-punch press thus has its primary use in the production of small batches of tablets, such as during formulation development, and during small-scale production, such as production for clinical trials.

Rotary press

The rotary press (also referred to as a multistation press) was developed to increase the output of tablets. The primary use of this machine is thus during scale-up in the latter part of the formulation work, and during large-scale production. Outputs of over 10 000 tablets per minute can be achieved by rotary presses.

A rotary press operates with a number of dies and sets of punches, which can vary considerably from three for small rotary presses up to 60 or more for large presses. The dies are mounted in a circle in the die table and both the die table and the punches rotate together during operation of the machine, so that one die is always associated with one pair of

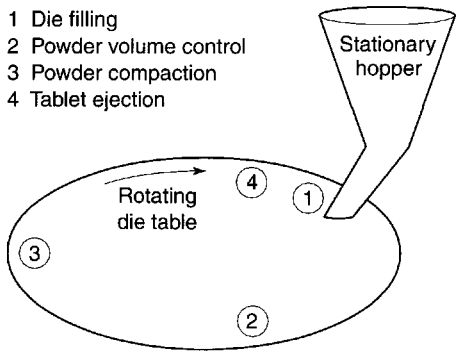


Fig. 27.3 Schematic illustration of the events involved in the formation of tablets with a rotary press.

punches (Figs. 27.3 and 27.4). The vertical movement of the punches is controlled by tracks that pass over cams and rolls used to control the volume of powder fed into the die and the pressure applied during compression.

The powder is held in a hopper whose lower opening is located just above the die table. The powder flows by gravity on to the die table and is fed into the die by a feed frame. The reproducibility of the die feeding can be improved by a rotating device, referred to as a force-feeding device. During powder compression both punches operate by vertical movement. After tablet ejection, the tablet is knocked away as the die passes the feed frame.

Computerized hydraulic press

For computerized hydraulic presses the movement of the punches can be controlled and varied considerably. Thus, tablets can be prepared under controlled conditions with respect to the loading pattern and loading rate. Possible applications are the investigation of the sensitivity of a drug to such variations, or to mimic the loading pattern of production presses to predict scale-up problems. Because of this latter application, this type of press is also referred to as a 'simulator'.

Instrumentation of tablet presses

Significant research on the process of tablet preparation was initiated in the 1940s and 1950s, i.e. about 100 years after the introduction of tablets as a dosage form. An important step in the development of such fundamental research was the introduction of instrumented tablet machines. By this instrumentation, the forces involved in the compaction process, i.e. the press forces from the upper and lower punches and

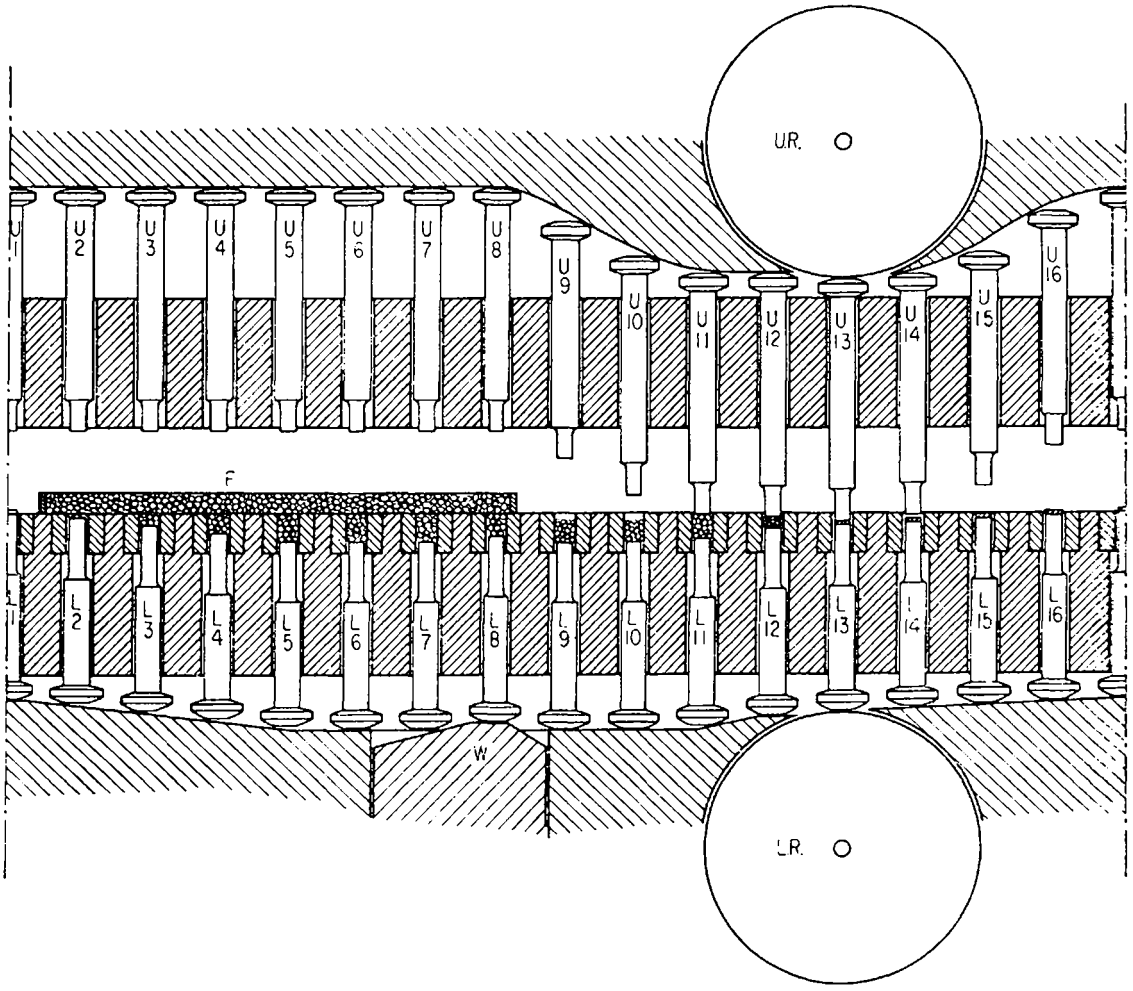


Fig. 27.4 Diagram of punch tracks of a rotary tablet press. UR, upper roller; LR, lower roller; W, powder volume adjuster; F, feed frame with granules. U1 to U8, upper punches in raised position; L1, lower punch at top position, tablet ejected; L2 to L7, lower punches dropping to lowest position and filling die with granules to an overfill at L7; L8, lower punch raised to expel excess granules giving correct volume; U9 to U12, upper punches lowering to enter die at U12; L13 and U13, upper and lower punches pass between rollers and granules are compacted to a tablet; U14 to U16, upper punch rising to top position; L14 to L16, lower punch rising to eject tablet.

the force transmitted to the die, and the displacement of the upper and lower punch during the compression and decompression phases, could be recorded.

Instrumented presses are used in research, in development and in the production of tablets. In research and development, instrumented machines are used to provide fundamental information on the mechanical and compaction properties of powders that should be used in tablet formulations. With this application, the work is normally carried out by instrumented single-punch presses or with instrumented hydraulic presses (compaction simulators). The two main applications for an instrumented press in research and development are:

1. to prepare tablets under defined conditions, e.g. in terms of applied force during compaction. These tablets are thereafter characterized by different procedures, such as imaging, surface area and tensile strength analysis;
2. to describe and analyse the compression properties of materials by studying punch forces and punch displacements during the compression and decompression phases. A series of different procedures exists involving, for example, the assessment of deformation behaviour of particles during compression and friction properties during ejection. Some of these are described below.

In production, instrumented production machines, i.e. rotary presses, are used to control the tabletting operation and to ensure that tablets of consistent quality are produced. Normally, only force signals are used on production machines and the variation in force signal during compression is followed as it reflects variations in tablet weight.

Force transducers commonly used in the instrumentation of tablet machines are of two types. The most common type is called a ***strain gauge***, which consists of wires through which an electric current is passed. The strain gauge is bonded to a punch or punch holder. During powder compression, a force is applied to the punches and they will temporarily deform. The magnitude of this deformation is dependent on the elastic modulus of the punches and the force applied. When the punch is deformed the wire of the strain gauge is also deformed, and the electrical resistance of the strain gauge will change. This change in electrical resistance can be recorded and calibrated in terms of a force signal. Another, less common, type of force transducer, employs piezoelectric crystals. These are devices which emit an electrical charge when loaded, the magnitude of which is proportional to the applied force.

Displacement transducers measure the distance which the punches travel during the compression and decompression processes. The most common type of displacement transducer delivers an analogue signal. It consists of a rod and some inductive elements mounted in a tube. When the rod moves within the tube, a signal is obtained which directly reflects the position of the rod. The movable rod is connected to the punch so that they move in parallel, i.e. the signal from the displacement transducer reflects the position of the punch. Digital displacement transducers are also used in instrumented tablet machines. Such transducers are based on differences in signal level depending on the position of an indicator. One advantage of a digital displacement transducer is that it is insensitive to electrical noise.

Displacement transducers are necessarily mounted some distance from the punch tip. There is therefore a difference in the position given by the transducer and the real position of the punch tip owing to deformation of the punch along the distance between its tip and the connection point of the transducer. This deviation must be determined by a calibration procedure, e.g. by compressing the punch tips against each other, and a correction for this error must be made before the displacement data can be used.

The signals from the force and displacement transducers are normally amplified and sampled into a computer. After conversion into digital form, the signals are transformed into physically relevant units, i.e. N, Pa, μm etc., and organized as a function of time. To obtain reliable data the calibration of the signals, the resolution of the measuring systems and the reproducibility of the values must be carefully considered.

Technical problems during tabletting

A number of technical problems can arise during the tabletting procedure, among which the most important are:

- high weight and dose variation of the tablets
- low mechanical strength of the tablets
- capping and lamination of the tablets
- adhesion or sticking of powder material to punch tips
- high friction during tablet ejection.

Such problems are related to the properties of the powder intended to be formed into tablets, and also to the design and conditions of the press. They should therefore be avoided by ensuring that the powder possesses adequate technical properties and also that a suitable, well conditioned tablet press is used, e.g. in terms of the use of forced-feed devices and polished and smooth dies and punches.

Important technical properties of a powder which must be controlled to ensure the success of a tabletting operation are:

- homogeneity and segregation tendency
- flowability
- compression properties and compactability
- friction and adhesion properties.

The technical properties of the powder are controlled by the ingredients of the formulation (i.e. the drug and excipients) and by the way by which the ingredients are combined into a powder during precompaction processing. The precompaction processing often consists of a series of unit operations in sequence. The starting point is normally the drug in a pure, most often crystalline form; the subsequent treatment of the drug particles is sometimes referred to as ***downstream processing***. The unit operations used during this precompaction treatment are mainly particle size reduction, powder mixing, particle size enlargement and powder drying. For further details see Chapters 11,13, 25 and 26, respectively. Traditionally, the use of a particle size enlargement operation, normally referred to as granulation, is the dominant procedure in preparing a powder for tabletting. To save time and energy, precompaction

processing without a particle size enlargement operation is chosen if possible. This procedure is called ***tablet production by direct compression***, or ***direct compaction***.

Tablet production via granulation

Rationale for granulating powders prior to tabletting

Because both granulation and tabletting involve the formation of aggregates, tablet production by granulation is based on the combination of two size enlargement processes in sequence. The main rationales for granulating the powder (drug and filler mixture) before tabletting are:

- to increase the bulk density of the powder mixture and thus ensure that the required volume of powder can be filled into the die;
- to improve the flowability of the powder in order to ensure that tablets with a low and acceptable tablet weight variation can be prepared;
- to improve mixing homogeneity and reduce segregation by mixing small particles which subsequently adhere to each other;
- to improve the compactability of the powder by adding a solution binder, which is effectively distributed on the particle surfaces;
- to ensure a homogenous colour in a tablet by adding the colour so it is distributed effectively over the particle surfaces;
- to affect the dissolution process for hydrophobic, poorly soluble particles by using a fine particulate drug which is thoroughly mixed with a hydrophilic filler and a hydrophilic binder.

Before granulation the drug might be processed separately in order to obtain a suitable quality in terms of solid-state and particulate properties, such as spray-drying and milling. Normally, the drug exists in dry particulate form before granulation. However, it might be suspended or dissolved in a liquid and be added to the filler as a part of the agglomeration liquid.

Different procedures may be used to prepare a granulation, among which the most important are the use of convective mixers, fluidized-bed driers, spray driers and compaction machines (see Chapter 25 for further detail).

Granulation by convective mixing

Agitation of a powder by convection in the presence of a liquid followed by drying, is the main procedure for the preparation of a pharmaceutical granulation. This is considered to be the most effective means in terms of production time and cost to prepare good-quality granulations. The process is often referred to as ***wet granulation***.

The ingredients to be granulated in a convective mixer are first dry mixed. The objective is to achieve a good homogeneity. As the components are often cohesive powders, a convective mixer operating at high intensity is normally used (a high-shear mixer). The mixture often consists of the drug and a filler. A disintegrant may also be included (i.e. an intragranular disintegrant), but it is also common to add the disintegrant to the dry granulation (i.e. an extragranular disintegrant). After wet mixing the wet mass is dried in a separate drier (a fluidized-bed dryer or a tray dryer). Because granulation in a convective mixer is not a very well controlled operation, large granules (above 1 mm) are often formed which must be broken down into smaller units. This is normally done by milling in a hammer mill or by pressing the granulation through a screen in an oscillating granulator. Granules ranging in size from about 100 to 800 μm are thus obtained.

The prepared granulation is finally dry-mixed with the other ingredients, for example in a double-cone mixer, before tabletting. Common excipients added in this final mixing operation are disintegrants, lubricants, glidants and colourants. Figure 27.5 summarizes the sequence of unit operations used in the production of tablets with precompaction treatment by granulation.

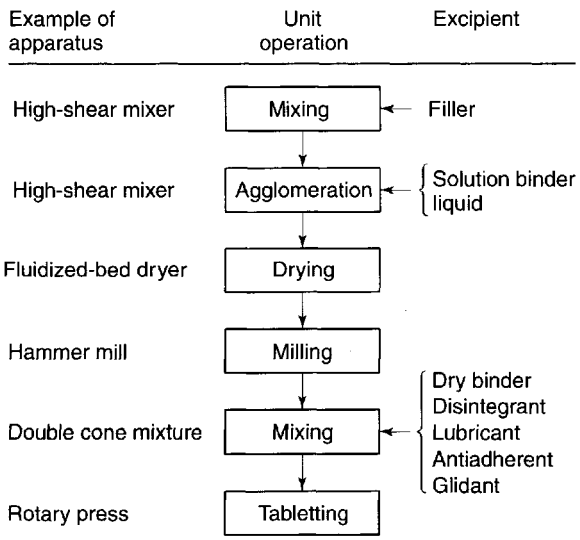


Fig. 27.5 *Overview of the sequence of unit operations used in the production of tablets with precompaction treatment by granulation.*

Alternative granulation procedures

A series of alternative granulation procedures can be preferable in certain situations. Granulation in a fluidized-bed apparatus is less common than the use of convective mixers as it is considered to be more time-consuming. However, granulations of high quality in terms of homogeneity, flowability and compactability can be prepared by this operation.

By spray drying a suspension of drug particles in a liquid, which can contain a dissolved binder, relatively small spherical granules with uniform size can be prepared. The process is of limited use except for the preparation of fillers or diluents for direct compaction. The granulation can show a good compactability and presents a possibility to granulate a drug suspension without a separate drying step for the drug substance.

The formation of granules by compacting the powder into large compacts which are subsequently comminuted into smaller granules (often referred to as ***dry granulation***, or ***slugging***) is a possible granulation procedure which, however, is not widely used in pharmaceutical production. The procedure can be employed as a means to avoid exposure of the powder to moisture and heat. In addition, for powders of very low bulk density compaction can be an effective means to increase markedly their bulk density.

Tablet production by direct compaction

An obvious way to reduce production time and hence cost is to minimize the number of operations involved in the pretreatment of the powder mixture before tabletting. Tablet production by direct compaction involves only two operations in sequence, powder mixing and tabletting (Fig. 27.6). The advantage with direct compaction is primarily a reduced production cost. However, in a direct compactable formulation specially designed fillers and dry binders are normally required, which usually are more expensive than the traditional ones. They may also require a larger number of quality tests before processing. As heat and water are not involved, product stability can be improved. Finally, drug dissolution might be faster from a tablet prepared by direct compaction owing to fast tablet disintegration into primary drug particles.

The disadvantages of direct compaction are mainly technological. In order to handle a powder of acceptable flowability and bulk density, relatively large particles must be used which, firstly, may be difficult to mix to a high homogeneity, and secondly are prone to segregate. Moreover, a powder consisting mainly of drug will be difficult to form into tablets if the drug itself has poor compactability. Finally, an even colouring of tablets can be difficult to achieve with a colourant in dry particulate form.

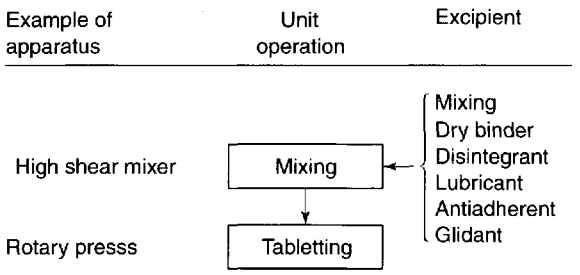


Fig. 27.6 Overview of the sequence of unit operations used in the production of tablets by direct compaction.

Direct compaction has been used mainly for two types of drug, firstly, relatively soluble drugs which can be processed as coarse particles (to ensure good flowability) and, secondly, relatively potent drugs which are present in a few milligrams in each tablet and can be mixed with relatively coarse excipient particles (in this latter case the flow and compaction properties of the formulation are controlled mainly by the excipients).

TABLET EXCIPIENTS

In addition to the active ingredient(s), a series of excipents are normally included in a tablet; their role is to ensure that the tabletting operation can run satisfactorily and to ensure that tablets of specified quality are prepared. Depending on the intended main function, excipients to be used in tablets are subcategorized into different groups. However, one excipient can affect the properties of a powder or the tablet in a series of ways, and many substances used in tablet formulations can thus be described as multifunctional. The functions of the most common types of excipients used in tablets are described below. Examples of substances used as excipients in tablets are given in Table 27.1.

Filler (or diluent)

In order to form tablets of a size suitable for handling, a lower limit in terms of powder volume and weight is required. Tablets weigh normally at least 50 mg. Therefore, a low dose of drug per tablet requires the incorporation of a substance into the formulation to increase the bulk volume of the powder and

Table 27.1 Examples of substances used as excipients in tablet formulation

Type of excipient	Example of substances
Filler	Lactose Sucrose Glucose Mannitol Sorbitol Calcium phosphate Calcium carbonate Cellulose
Disintegrant	Starch Cellulose Crosslinked polyvinyl pyrrolidone Sodium starch glycolate Sodium carboxymethyl cellulose
Solution binder	Gelatin Polyvinyl pyrrolidone Cellulose derivatives (e.g. hydroxypropylmethyl cellulose) Polyethylene glycol Sucrose Starch
Dry binder	Cellulose Methyl cellulose Polyvinyl pyrrolidone Polyethylene glycol
Glidant	Silica Magnesium stearate Talc
Lubricant	Magnesium stearate Stearic acid Polyethylene glycol Sodium lauryl sulphate Sodium stearyl fumarate Liquid paraffin
Antiadherent	Magnesium stearate Talc Starch Cellulose

hence the size of the tablet. This excipient, known as the filler or the diluent, is not necessary if the dose of the drug per tablet is high.

The ideal filler should fulfil a series of requirements, such as:

- be chemically inert
- be non-hygroscopic
- be biocompatible
- possess good biopharmaceutical properties (e.g. water soluble or hydrophilic)
- possess good technical properties (such as compactability and dilution capacity)
- have an acceptable taste
- be cheap.

As all these requirements cannot be fulfilled by a single substance, different substances have gained use as fillers in tablets, mainly carbohydrates but also some inorganic salts.

Lactose is the most common filler in tablets. It possesses a series of good filler properties, e.g. dissolves readily in water, has a pleasant taste, is non-hygroscopic and fairly non-reactive and shows good compactability. Its main limitation is that some people have an intolerance to lactose.

Lactose exists in both crystalline and amorphous form. Crystalline lactose is formed by precipitation and, depending on the crystallization conditions, α-monohydrate or β-lactose (an anhydrous form) can be formed. By thermal treatment of the monohydrate form, crystalline α-anhydrous particles can be prepared. Depending on the crystallization conditions and the use of subsequent size reduction by milling, lactoses of different particle sizes is obtained.

Amorphous lactose can be prepared by spray-drying a lactose solution (giving nearly completely amorphous particles) or a suspension of crystalline lactose particles in a lactose solution (giving aggregates of crystalline and amorphous lactose). Amorphous lactose dissolves more rapidly than crystalline and shows better compactability. Its main use is therefore in the production of tablets by direct compaction. The amorphous lactose is, however, hygroscopic and physically unstable, i.e. it will spontaneously crystallize if crystallization conditions are met as a result of elevated temperature or high relative humidity.

Other sugars or sugar alcohols, such as glucose, sucrose, sorbitol and mannitol, have been used as alternative fillers to lactose, primarily in lozenges or chewable tablets because of their pleasant taste. Mannitol has a negative heat of solution and imparts a cooling sensation when sucked or chewed.

Apart from the sugars perhaps the most widely used fillers are celluloses in powder forms of different types. Celluloses are biocompatible, chemically inert and have good tablet-forming and disintegrating properties. They are therefore used also as dry binders and disintegrants in tablets. They are compatible with many drugs but, owing to their hygroscopicity, may be incompatible with drugs prone to hydrolyse in the solid state.

The most common type of cellulose powder used in tablet formulation is microcrystalline cellulose. The name indicates that the particles have both crystalline and amorphous regions, depending on the relative position of the cellulose chains within the solid. The crystallinity might vary depending on the source of the cellulose and the preparation

procedure. The degree of crystallinity will affect the physical and technical properties of the particles, e.g. in terms of hygroscopicity and powder compactability.

Microcrystalline cellulose is prepared by hydrolysis of cellulose followed by spray drying. The particles thus formed are aggregates of smaller cellulose fibres. Depending on the preparation conditions, aggregates of different particle size can be prepared which have different flowabilities.

A final important example of a common filler is an inorganic substance, dicalcium phosphate dihydrate. This is insoluble in water and non-hygroscopic but is hydrophilic, i.e. easily wetted by water. The substance can be obtained both in a fine particulate form, mainly used in granulation, and in an aggregated form. The latter possesses good flowability and is used in tablet production by direct compaction. Calcium phosphate is slightly alkaline and may thus be incompatible with drugs sensitive to alkaline conditions.

Disintegrant

A disintegrant is included in the formulation to ensure that the tablet, when in contact with a liquid, breaks up into small fragments, which promotes rapid drug dissolution. Ideally, the tablet should break up into individual drug particles in order to obtain the largest possible effective surface area during dissolution.

The disintegration process for a tablet occurs in two steps. First, the liquid wets the solid and penetrates the pores of the tablet. Thereafter, the tablet breaks into smaller fragments. The actual fragmentation of the tablet can also occur in steps, i.e. the tablet disintegrates into aggregates of primary particles which subsequently deaggregate into their primary drug particles. A deaggregation directly into primary powder particles will set up conditions for the fastest possible dissolution of the drug. A scheme for the release of the drug from a disintegrating tablet is shown in Fig. 27.7.

Several mechanisms of action of disintegrants have been suggested, such as swelling of particles, exothermic wetting reaction, particle repulsion and particle deformation recovery. However, as two main processes are involved in the disintegration event, disintegrants to be used in plain tablets are here classified into two types:

1. *Disintegrants that facilitate water uptake.* These disintegrants act by facilitating the transport of liquids into the pores of the tablet, with the consequence that the tablet may break into fragments. One obvious type of substance that can promote liquid penetration are surface active agents. Such substances are used to make the drug particle surfaces more hydrophilic and thus promote the wetting of the solid and the penetration of the liquid into the pores of the tablet. It has also been suggested that other substances can promote the liquid penetration using capillary forces to suck water into the pores of the tablet.
2. *Disintegrants that will rupture the tablet.* Rupturing of tablets can be caused by swelling of the disintegrant particles during sorption of water. However, it has also been suggested that non-swelling disintegrants can break the tablet, and different mechanisms have been suggested. One such concerns a repulsion of particles in contact with water and another the recovery of deformed particles to their original shape in contact with water, i.e. particles which have been deformed during compaction.

The most traditionally used disintegrant in conventional tablets is starch, among which potato, maize and corn starches are the most common types used. The typical concentration range of starch in a tablet formulation is up to 10%. Starch particles swell in contact with water and this swelling can subsequently disrupt the tablet. However, it has also been suggested that starch particles may facilitate disintegration by particle–particle repulsion.

The most common and effective disintegrants act via a swelling mechanism and a series of effective swelling disintegrants have been developed which can swell dramatically during water uptake and thus quickly and effectively break the tablet. These are normally modified starch or modified cellulose. High-swelling disintegrants are included in the formulation at relatively low concentrations, typically 1–5% by weight.

Disintegrants can be mixed with other ingredients prior to granulation and thus be incorporated within the granules (intragranular addition). It is also common for the disintegrant to be mixed with the dry granules before the complete powder mix is compacted (extragranular addition). The latter procedure will contribute to an effective disintegration of the tablet into smaller fragments. Disintegrants may also be incorporated as both an intragranular and an extragranular portion.

A third group of disintegrants functions by producing gas, normally carbon dioxide, in contact with water. Such disintegrants are used in effervescent tablets and normally not in tablets that should be swallowed as a solid. The liberation of carbon dioxide is

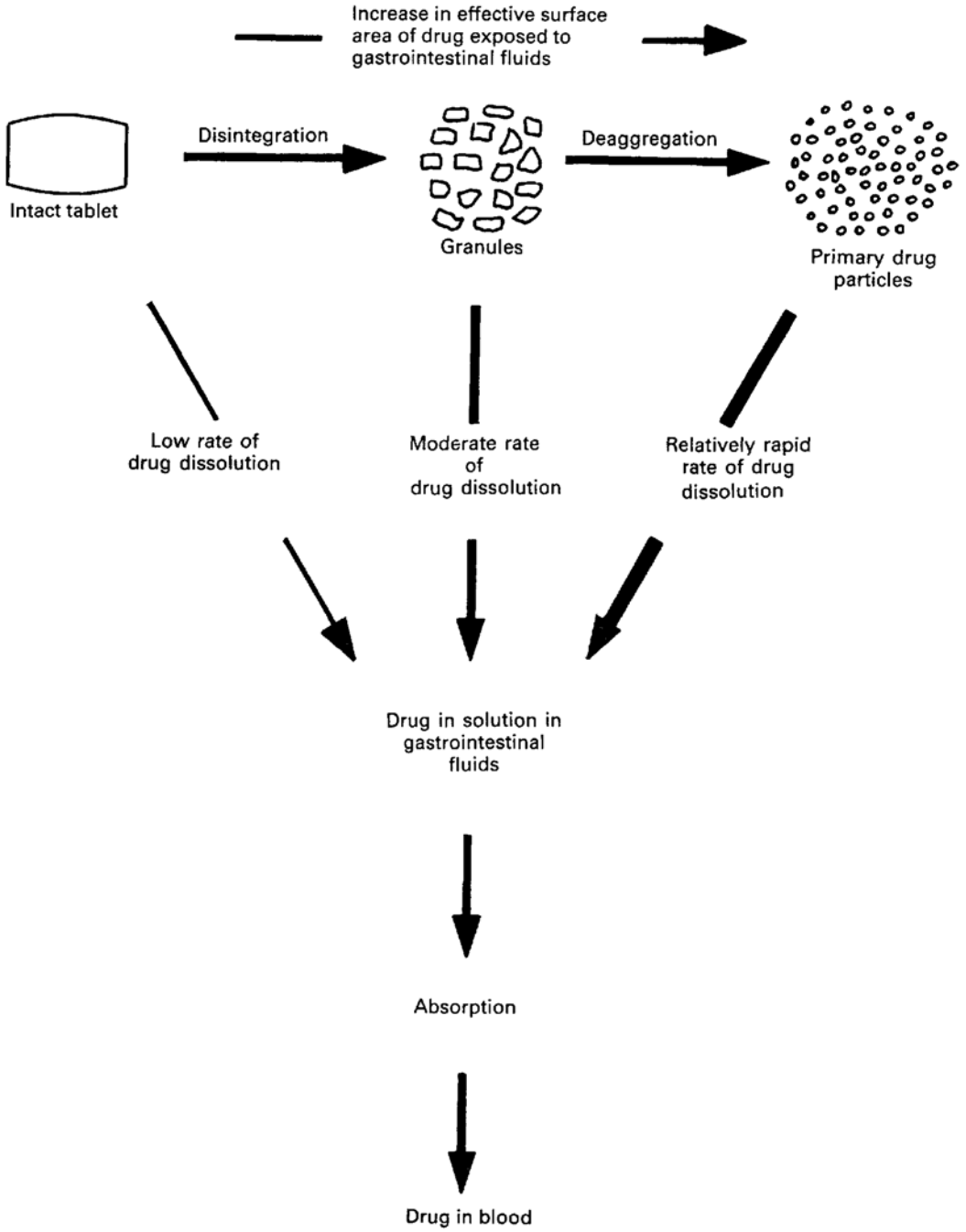


Fig. 27.7 Mechanistic representation of the drug release process from a tablet by disintegration and dissolution. From Wells, J.I. and Rubinstein, M.W. (1976) *Pharm. J.* **217**, 629.

obtained by the decomposition of bicarbonate or carbonate salts in contact with acidic water. The acidic pH is accomplished by the incorporation of a weak acid in the formulation, such as citric acid and tartaric acid.

Binder

A binder (also sometimes called adhesive) is added to a drug–filler mixture to ensure that granules and tablets can be formed with the required mechanical

strength. Binders can be added to a powder in different ways:

- As a dry powder which is mixed with the other ingredients before wet agglomeration. During the agglomeration procedure the binder might thus dissolve partly or completely in the agglomeration liquid;
- As a solution which is used as agglomeration liquid during wet agglomeration. The binder is here often referred to as a ***solution binder***.
- As a dry powder which is mixed with the other ingredients before compaction (slugging or tabletting). The binder is here often referred to as a ***dry binder***.

Both solution binders and dry binders are included in the formulation at relatively low concentrations, typically 2–10% by weight. Common traditional solution binders are starch, sucrose and gelatin. More commonly used binders today, with improved adhesive properties, are polymers such as polyvinylpyrrolidone and cellulose derivatives (in particular hydroxypropyl methylcellulose). Important examples of dry binders are microcrystalline cellulose and crosslinked polyvinylpyrrolidone.

Solution binders are generally considered the most effective, and this is therefore the most common way of incorporating a binder into granules; the granules thus formed are often referred to as binder–substrate granules. It is not uncommon, however, for a dry binder to be added to the dry binder–substrate granules before tabletting in order to further improve the compactability of the granulation.

Glidant

The role of the glidant is to improve the flowability of the powder. This is especially important during tablet production at high production speeds and during direct compaction. However, because the requirement for adequate flow is high, a glidant is often also added to a granulation before tabletting.

Traditionally, talc has been used as a glidant in tablet formulations, in concentrations of about 1–2% by weight. Today, the most commonly used glidant is probably colloidal silica, added in very low proportions (about 0.2% by weight). Because the silica particles are very small they adhere to the particle surfaces of the other ingredients and improve flow by reducing interparticulate friction. Magnesium stearate, normally used as a lubricant, can also promote powder flow at low concentrations (< 1% by weight).

Lubricant

The function of the lubricant is to ensure that tablet formation and ejection can occur with low friction between the solid and the die wall. High friction during tabletting can cause a series of problems, including inadequate tablet quality (capping or even fragmentation of tablets during ejection, and vertical scratches on tablet edges) and may even stop production. Lubricants are thus included in almost all tablet formulations.

Lubrication is achieved by mainly two mechanisms: ***fluid lubrication*** and ***boundary lubrication*** (Fig. 27.8). In fluid lubrication a layer of fluid is located between and separates the moving surfaces of the solids from each other and thus reduces the friction. Fluid lubricants are seldom used in tablet formulations. However, liquid paraffin has been used, such as in formulations for effervescent tablets.

Boundary lubrication is considered as a surface phenomenon, as here the sliding surfaces are separated by only a very thin film of lubricant. The nature of the solid surfaces will therefore affect friction. In boundary lubrication the friction coefficient and wear of the solids are higher than with fluid lubrication. All substances that can affect interaction between sliding surfaces can be described as boundary lubricants, including adsorbed gases. The lubricants used in tablet formulations acting by boundary lubrication are fine particulate solids.

A number of mechanisms have been discussed for these boundary lubricants, including that lubricants are substances that show a low resistance towards shearing. The most effective of the boundary lubricants are stearic acid or stearic acid salts, primarily

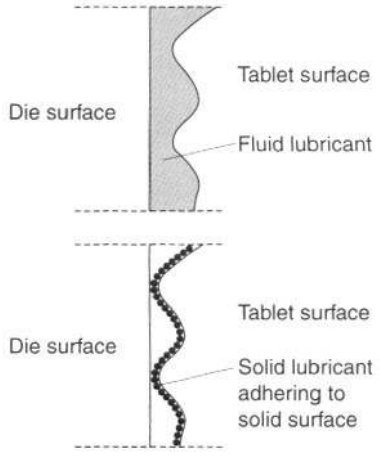


Fig. 27.8 Schematic illustration of lubrication mechanisms by fluid and boundary lubrication.

magnesium stearate. Magnesium stearate has become the most widely used lubricant owing to its superior lubrication properties. The stearic acid salts are normally used at low concentrations (< 1% by weight).

Besides reducing friction lubricants may cause undesirable changes in the properties of the tablet. The presence of a lubricant in a powder is thought to interfere in a deleterious way with the bonding between the particles during compaction, and thus reduce tablet strength (Fig. 27.9). Because many lubricants are hydrophobic, tablet disintegration and dissolution are often retarded by the addition of a lubricant. These negative effects are strongly related to the amount of lubricant present, and a minimum amount is normally used in a formulation, i.e. concentrations of 1% or below. In addition, the way in which the lubricant is mixed with the other ingredients should also be considered. It can, for example, be important if the excipients are added sequentially to a granulation rather than simultaneously. The total mixing time and the mixing intensity are also important in this context.

The commonly observed retardation of disintegration and dissolution of tablets is related to the hydrophobic character of the most commonly used lubricants. In order to avoid these negative effects more hydrophilic substances have been suggested as alternatives to the hydrophobic lubricants. Examples are surface-active agents and polyethylene glycol. A combination of hydrophobic and hydrophilic substances might also be used.

Both the effect on friction and the effect on the changes in tablet properties of a lubricant are related to the tendency of lubricants to adhere to the surface of drugs and fillers during dry mixing. Lubricants are often fine particulate substances which thus are prone to adhere to larger particles. In addition, studies on the mixing behaviour of magnesium stearate have indicated that this substance has the ability to form a film which can cover a fraction of the surface area of the drug or filler particles (the substrate particles). This film can be described as being continuous rather than particulate. A number of factors have been suggested to affect the development of such a lubricant film during mixing, and hence also affect friction and changes in tablet properties, such as the shape and surface roughness of the substrate particles; the surface area of the lubricant particles; mixing time and intensity; and the type and size of mixer.

Concerning the tablet strength-reducing effect of a lubricant, apart from the degree of surface coverage of the lubricant film obtained during mixing, the compression behaviour of the substrate particles will also be important. Drugs and fillers can thus be evaluated in terms of their lubricant sensitivity, i.e. the reduction in tablet strength due to the addition of a lubricant compared to a tablet formed from a powder without a lubricant. An important property for this lubricant sensitivity seems to be the degree of fragmentation the substrate particles undergo during compression (see below). It is thus assumed that, during compression,

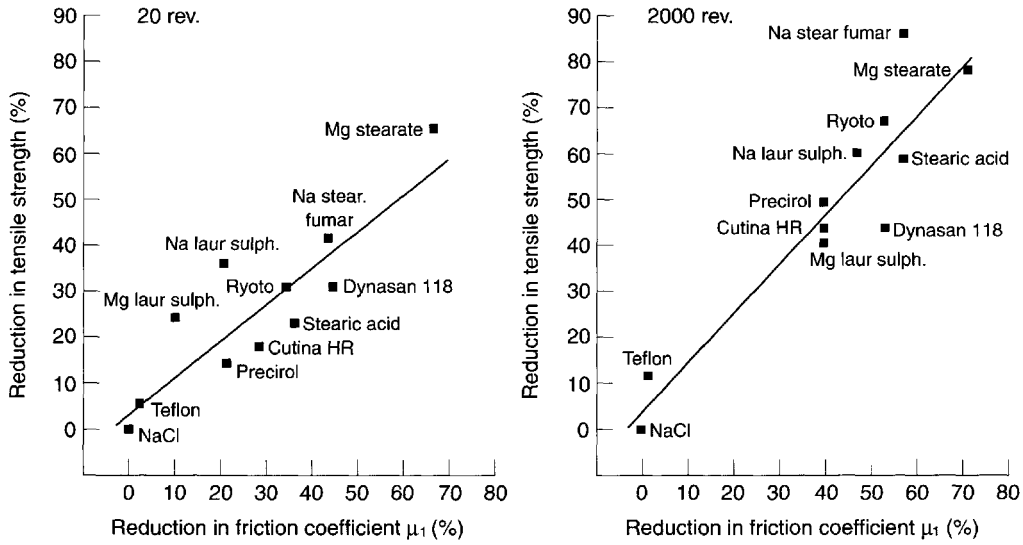


Fig. 27.9 The reduction in tablet tensile strength as a function of the reduction in friction coefficient during tabletting of a sodium chloride powder mixed with 0.1% by weight of a series of lubricants admixed at two different mixing intensities. (From Hölzer, A.H. and Sjögren, J. Acta Pharm. Suec. 18, 139, 1981, with permission).

particle surfaces which are not covered with a lubricant film are formed during particle fragmentation, and that these clean surfaces will bond differently from the lubricant-covered particle surfaces.

To explain the effect of lubricant film formation on the tensile strength of tablets, a coherent matrix model has been developed. This suggests that when a continuous matrix of lubricant-covered particle surfaces exists in a tablet, along which a fracture plane can be formed, the tablet strength is considerably lower than that of tablets formed from unlubricated powder. However, if the mixing and compression processes do not result in such a coherent lubricant matrix within the tablet, for example due to irregular substrate particles or particle fragmentation, the lubricant sensitivity appears to be lower.

Antiadherent

The function of an antiadherent is to reduce adhesion between the powder and the punch faces and thus prevent particles sticking to the punches. Many powders are prone to adhere to the punches, a phenomenon (known in the industry as ***sticking*** or ***picking***) which is affected by the moisture content of the powder. Such adherence is especially prone to happen if the tablet punches are engraved or embossed. Adherence can lead to a build-up of a thin layer of powder on the punches, which in turn will lead to an uneven and matt tablet surface with unclear engravings.

Many lubricants, such as magnesium stearate, have also antiadherent properties. However, other substances with limited ability to reduce friction can also act as antiadherents, such as talc and starch.

Sorbent

Sorbents are substances that are capable of sorbing some quantities of fluids in an apparently dry state. Thus, oils or oil–drug solutions can be incorporated into a powder mixture which is granulated and compacted into tablets. Microcrystalline cellulose and silica are examples of sorbing substances used in tablets.

Flavour

Flavouring agents are incorporated into a formulation to give the tablet a more pleasant taste or to mask an unpleasant one. The latter can be achieved also by coating the tablet or the drug particles.

Flavouring agents are often thermolabile and so cannot be added prior to an operation involving heat. They are often mixed with the granules as an alcohol solution.

Colourant

Colourants are added to tablets to aid identification and patient compliance. Colouring is often accomplished during coating (see Chapter 28 for further information), but a colourant can also be included in the formulation prior to compaction. In the latter case the colourant can be added as an insoluble powder or dissolved in the granulation liquid. The latter procedure may lead to a colour variation in the tablet caused by migration of the soluble dye during the drying stage (see Chapter 26 for more information on the phenomenon of solute migration).

TABLET TYPES

Classification of tablets

Based on their drug-release characteristics, tablets can be classified into three types, immediate release, extended release and, delayed release. For immediate-release tablets the drug is intended to be released rapidly after administration, or the tablet is dissolved and administered as a solution. This is the most common type of tablet and includes disintegrating, chewable, effervescent, sublingual and buccal tablets.

Modified-release tablets should normally be swallowed intact. The formulation and thus also the type of excipients used in such tablets might be quite different from those of immediate-release tablets. The drug is released from an extended-release tablet slowly at a nearly constant rate. If the rate of release is constant during a substantial period of time, a zero-order type of release is obtained, i.e. $M = kt$ (where M is the cumulative amount of drug released and t is the release time). This is sometimes described as an ideal type of extended-release preparation. However, for most type of extended-release tablets a perfect zero-order release is not obtained.

For delayed-release tablets the drug is liberated from the tablet some time after administration. After this period has elapsed, the release is normally rapid. The most common type of delayed-release tablet is an enteric tablet, for which the drug is released in the upper part of the small intestine after the preparation has passed the stomach. However, a delayed-release can also be combined with a slow drug release, e.g. for local treatment in the lower part of the intestine or in the colon.

The type of release obtained from immediate-, extended- and delayed-release tablets is illustrated in Figure 27.10.

Disintegrating tablets

The most common type of tablet is intended to be swallowed and to release the drug in a relatively short time thereafter by disintegration and dissolution, i.e. the goal of the formulation is fast and complete drug release in vivo. Such tablets are often referred to as conventional or plain tablets. A disintegrating tablet includes normally at least the following type of excipients: filler (if the dose of the drug is low), disintegrant, binder, glidant, lubricant and antiadherent.

As discussed above, the drug is released from a disintegrating tablet in a sequence of processes, including tablet disintegration, drug dissolution and drug absorption (see Fig. 27.7). All these processes will affect, and can be rate-limiting steps for, the rate of drug bioavailability. The rate of the processes is affected by both formulation factors and production conditions.

The disintegration time of the tablet can be markedly affected by the choice of excipients, especially disintegrant (Fig. 27.11). The type of filler and lubricant can also be of significant importance for tablet disintegration.

Tablet disintegration may also be affected by production conditions during manufacture. Important examples are the design of the granulation procedure (which will affect the physical properties of the granules), mixing conditions during the addition of lubricants and antiadherents, and the applied punch force during tabletting and the punch force–time relationship. It has been reported that an increased compaction pressure can either increase or decrease disintegration time, or give complex relationships with maximum or minimum disintegration times.

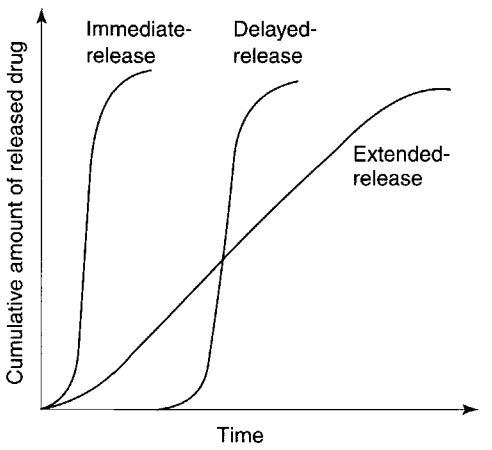


Fig. 27.10 Schematic representation of the cumulative amount of drug released from immediate-, extended- and delayed-release tablets.

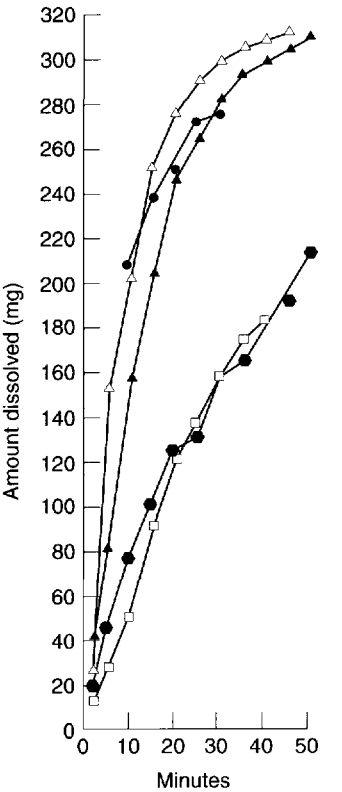


Fig. 27.11 The dissolution rate of salicylic acid, as assessed by an in vitro dissolution method based on agitated baskets, from tablets formed from mixtures of salicylic acid (325 mg) and a series of different types of starches as disintegrant. (From Underwood, T.W. and Cadwallader, D.E., J. Pharm. Sci. 61, 239, 1972.) □ potato starch, ● arrowroot starch, ▲ rice starch, • corn starch, △ compressible starch.

For poorly water-soluble drugs the dissolution rate is often the rate-limiting step for bioavailability. The dissolution rate is a function of the solubility and the surface area of the drug (see Chapter 2). Thus, dissolution rate will increase if the solubility of the drug is increased, e.g. by the use of a salt of the drug. It is also possible to speed up the dissolution process by incorporating into the formulation a substance that forms a salt with the drug during dissolution. This has been a common means to increase the dissolution rate of aspirin by using magnesium oxide in the formulation.

The drug dissolution rate will also increase with an increase in the surface area of the drug. Thus,

control of drug particle size is important to control drug dissolution. However, a reduced particle size will make a powder more cohesive. A reduction in drug particle size might thus give aggregates of particles which are difficult to break up, with the consequence that the drug dissolution rate from the tablet will be reduced. It is thus important to ensure that the tablet is formulated in such a way that it will disintegrate, and the aggregates thus formed break up into small drug particles so that a large surface area of the drug is exposed to the dissolution medium.

For drugs with poor absorption properties the absorption can be affected (see Chapter 17) by modifying the drug lipophilicity, e.g. by esterification of the drug. The use of substances in the formulation that affect the permeability of the gastrointestinal cell membranes, often referred to as absorption enhancers, is also a possible means to increase the drug absorption rate and degree.

Single disintegrating tablets can also be prepared in the form of multilayers, i.e. the tablet consists of two or three layers cohered to each other (double- and triple-layered tablets). During the preparation of multilayer tablets the die is filled in two or three consecutive steps with different granulations from separate feed stations. Each layer is normally compressed after each fill.

Multilayer tablets are made primarily to separate incompatible drugs from each other, i.e. incompatible drugs can be incorporated into the same tablet. Although intimate contact exists at the surface between the layers, the reaction between the incompatible drugs is limited. The use of layered tablets where the layers are differently coloured represents an approach to preparing easily identifiable tablets.

Another variation of the disintegrating tablet is coated tablets which are intended to disintegrate and release the drug quickly (in contrast to coated tablets intended for modified release). The rationale for using coated tablets and detailed descriptions of the procedures used for tablet coating (sugar coating, film coating and press coating) are given in Chapter 28.

Chewable tablets

Chewable tablets are chewed and thus mechanically disintegrated in the mouth. The drug is, however, normally not dissolved in the mouth but swallowed and dissolves in the stomach or intestine. Thus, chewable tablets are used primarily to accomplish a quick and complete disintegration of the tablet – and hence obtain a rapid drug effect – or to facilitate the intake of the tablet. A common example of the former is antacid tablets. In the latter case, the elderly and children in particular have difficulty in swallowing tablets, and so chewable tablets are attractive forms of medication. Important examples are vitamin tablets. Another general advantage of a chewable tablet is that this type of medication can be taken when water is not available.

Chewable tablets are similar in composition to conventional tablets except that a disintegrant is normally not included in the composition. Flavouring and colouring agents are common, and sorbitol and mannitol are common examples of fillers.

Effervescent tablets

Effervescent tablets are dropped into a glass of water before administration, during which carbon dioxide is liberated. This facilitates tablet disintegration and drug dissolution; the dissolution of the tablet should be complete within a few minutes. As mentioned above, the effervescent carbon dioxide is created by a reaction in water between a carbonate or bicarbonate and a weak acid such as citric or tartaric.

Effervescent tablets are used to obtain rapid drug action, for example for analgesic drugs (Fig. 27.12), or to facilitate the intake of the drug, for example for vitamins.

The amount of sodium bicarbonate in an effervescent tablet is often quite high (about 1 g). After dissolution of such a tablet, a buffered water solution will be obtained which normally temporarily increases the pH of the stomach. The result is a rapid emptying of the stomach and the residence time of the drug in the stomach will thus be short. As drugs

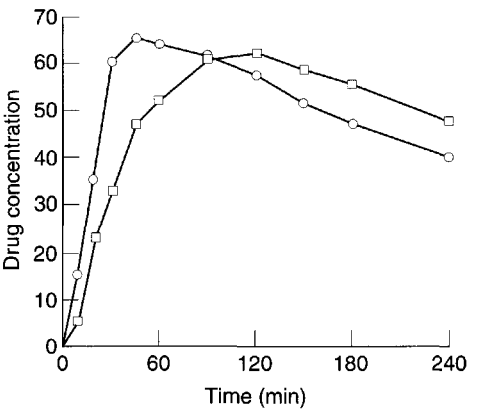


Fig. 27.12 Concentration of salicylates in plasma after administration of acetylsalicylic acid tablets (1 g). Circles, effervescent tablet; squares, conventional tablet. (From Ekenved, G., Elofsson, R. and Sölvell, L. Acta Pharm. Suec. 12, 323, 1975.)

are absorbed more effectively in the small intestine than in the stomach, effervescent tablets can thus show a fast drug bioavailability, which can be advantageous, for example, for analgesic drugs. Another aspect of the short residence time of the drug in the stomach is that drug-induced gastric irritation can be avoided, e.g. for aspirin tablets, as the absorption of aspirin in the stomach can cause irritation.

Effervescent tablets also often include a flavour and a colourant. A water-soluble lubricant is preferable in order to avoid a film of a hydrophobic lubricant on the surface of the water after tablet dissolution. A binder is normally not included in the composition.

Effervescent tablets are prepared by both direct compaction and by compaction via granulation. In the latter case, traditional wet granulation is seldom used; instead, granules are formed by the fusion of particles as a result of their partial dissolution during wet massing of a moistened powder.

Effervescent tablets should be packaged in such a way that they are protected against moisture. This is accomplished with waterproof containers, often including a dessicant, or with blister packs or aluminium foils.

Lozenges

Lozenges are tablets that dissolve slowly in the mouth and so release the drug dissolved in the saliva. Lozenges are used for local medication in the mouth or throat, e.g. with local anaesthesia, antiseptic and antibiotic drugs. They can thus be described as slow-release tablets for local drug treatment.

Disintegrants are not used in the formulation, but otherwise such tablets are similar in composition to conventional tablets. In addition, lozenges are often coloured and include a flavour. The choice of filler and binder is of particular importance in the formulation of lozenges, as these excipients should contribute to a pleasant taste or feeling during tablet dissolution. The filler and binder should therefore be water soluble and have a good taste. Common examples of fillers are glucose, sorbitol and mannitol. A common binder in lozenges is gelatin.

Lozenges are normally prepared by compaction at high applied pressures in order to obtain a tablet of high mechanical strength and low porosity which can dissolve slowly in the mouth.

Sublingual and buccal tablets

Sublingual and buccal tablets are used for drug release in the mouth followed by systemic uptake of the drug. A rapid systemic drug effect can thus be obtained without first-pass liver metabolism. Sublingual tablets are placed under the tongue and buccal tablets are placed in the side of the cheek.

Sublingual and buccal tablets are often small and porous, the latter facilitating fast disintegration and drug release.

Extended-release tablets

Classification of extended-release tablets

In recent years there has been great interest in the development and use of tablets which should be swallowed and thereafter slowly release the drug in the gastrointestinal tract. Such tablets are denominated in various ways, such as slow release, prolonged release, sustained release and extended-release. In the European Pharmacopoeia the term extended-release has been chosen as denominator for these types of tablets and so is used here. Extended-release tablets are often referred to as controlled-release preparations. This latter term is somewhat misleading, as all tablets, irrespective of their formulation and use, should release the drug in a controlled and reproducible way. (The nomenclature for extended-release preparations is subject to some debate and no worldwide acceptable system exists. The reader is referred to Chapter 20 for further discussion on this subject.)

After the release of the drug from the tablet the drug should normally be absorbed into the systemic circulation. The aim is normally to increase the time period during which a therapeutic drug concentration level in the blood is maintained. However, the aim can also be to increase the release time for drugs that can cause local irritation in the stomach or intestine if they are released quickly. Examples of the latter are potassium chloride and iron salts. In addition, drugs for local treatment of diseases in the large intestine are sometimes formulated as extended-release tablets.

An extended-release tablet contains one dose of the drug which is released for a period of about 12–24 hours. The release pattern can vary, from being nearly continuous to two or more pulses. In the latter case the pulses can correspond to a rapid release of the drug, or can be a combination of a rapid release of one portion of drug followed by a slow release of a second portion.

An extended-release preparation can also be categorized as a single-unit or a multiple-unit dosage form. In the first case the drug dose is incorporated into a single-release unit, and in the latter is divided into a large number of small release units. A multi-

ple-unit dosage form is often considered to give a more reproducible drug action.

There are a series of rationales behind the increased interest in administering drugs orally for systemic uptake in the form of extended-release tablets. However, the drug must fulfil certain criteria in order to render itself suitable for sustained-release medication, otherwise another type of tablet is a more feasible alternative. These rationales and criteria, as well as the pharmacokinetic aspects of extended-release drug administration, are described elsewhere in this book (Chapters 19 and 20). In Chapter 20 the formulation principles used to achieve extended drug release are described.

Extended-release tablets are often classified according to the mechanism of drug release. The following are the most common means used to achieve a slow, controlled release of the drug from tablets:

- Drug transport control by diffusion
- Dissolution control
- Erosion control
- Drug transport control by convective flow (accomplished by, for example, osmotic pumping)
- Ion-exchange control.

Diffusion-controlled release systems

In diffusion-controlled extended-release systems the transport by diffusion of dissolved drugs in pores filled with gastric or intestinal juice or in a solid (normally polymer) phase is the release-controlling process. Depending on the part of the release unit in which the drug diffusion takes place, diffusion-controlled release systems are divided into matrix systems (also referred to as monolithic systems) and reservoir systems. The release unit can be a tablet or a nearly spherical particle of about 1 mm in diameter (a granule or a millisphere). In both cases the release unit should stay more or less intact during the course of the release process. In matrix systems diffusion occurs in pores located within the bulk of the release unit, and in reservoir systems diffusion takes place in a thin water-insoluble film or membrane, often about 5–20 μm thick, which surrounds the release unit. Diffusion through the membrane can occur in pores filled with fluid, or in the solid phase that forms the membrane.

Drug is released from a diffusion-controlled release unit in two steps:

1. The liquid that surrounds the dosage form penetrates the release unit and dissolves the drug. A concentration gradient of dissolved drug is thus established between the interior and the exterior of the release unit.
2. The dissolved drug will diffuse in the pores of the release unit or the surrounding membrane and thus be released, or, alternatively, the dissolved drug will partition into the membrane surrounding the dose unit and diffuse in the membrane.

A dissolution step is thus normally involved in the release process, but the diffusion step is the rate-controlling step. The rate at which diffusion will occur depends on four variables: the concentration gradient over the diffusion distance, the area and distance over which diffusion occurs; and the diffusion coefficient of the drug in the diffusion medium. Some of these variables are used to modulate the release rate in the formulation.

Reservoir systems In a reservoir system the diffusion occurs in a thin film surrounding the release unit (Fig. 27.13). This film is normally formed from a high molecular weight polymer. The diffusion distance will be constant during the course of the release and, as long as a constant drug concentration gradient is maintained, the release rate will be constant, i.e. a zero-order release ($M = kt$).

One possible process for the release of the drug from a reservoir system involves partition of the drug dissolved inside the release unit to the solid membrane, followed by transport by diffusion of the drug within the membrane. Finally, the drug will partition to the solution surrounding the release unit. The driving force for the release is the concentration gradient of dissolved drug over the membrane. The release rate can be described in a simplified way by the following equation, which also summarizes the formulation factors by which the release rate can be controlled, i.e.

$$M/t = C\,A\,K\,D/h \qquad (27.1)$$

where C is the solubility of the drug in the liquid, A and h are the area and thickness of the membrane, D is the diffusion coefficient of the drug in the

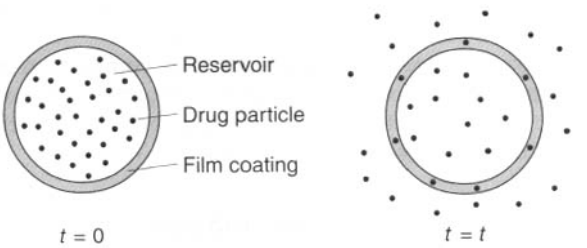


Fig. 27.13 Schematic illustration of the mechanism of drug release from a diffusion-based reservoir tablet (t = time).

membrane and K the partition coefficient for the drug between the membrane and the liquid at equilibrium.

In practice, the membrane surrounding the release unit often includes a water-soluble component. This can be small particles of a soluble substance, such as sucrose, or a water-soluble polymer, such as a water-soluble cellulose derivative (e.g. hydroxypropyl methylcellulose). In the latter case the polymer is used together with a water-insoluble polymer as the film-forming materials that constitute the coating. In such a membrane the water-soluble component will dissolve and form pores filled with liquid in which the drug can thereafter diffuse. The area and length of these pores will thus constitute the diffusion area and distance. These factors can be estimated from the porosity of the membrane (E) and the tortuosity (τ) of the pores (the tortuosity refers to the ratio between the actual transport distance in the pores between two positions and the transport distance in a solution). The release rate can thus be described in a simplified way as follows:

$$M/t = C\,A\,E\,D/h\,\tau \qquad (27.2)$$

The membrane porosity and pore tortuosity can be affected by the addition of water-soluble components to the membrane.

For oral preparations the film surrounding the release units is normally based on high molecular weight, water-insoluble polymers, such as certain cellulose derivatives (e.g. ethyl cellulose) and acrylates. The film often also includes a plasticizer. In the case of drug release through liquid-filled pores a small amount of a water-soluble compound is also added, as described above. Reservoir systems today are normally designed as multiple-unit systems rather than single units.

Matrix systems In a matrix system the drug is dispersed as solid particles within a porous matrix formed of a water-insoluble polymer, such as polyvinyl chloride (Fig. 27.14). Initially, drug particles located at the surface of the release unit will be dissolved and the drug released rapidly. Thereafter, drug particles at successively increasing distances from the surface of the release unit will be dissolved and released by diffusion in the pores to the exterior of the release unit. Thus, the diffusion distance of dissolved drug will increase as the release process proceeds. The drug release, in terms of the cumulative amount of drug (M) released from a matrix in which drug particles are suspended is proportional to the square root of time i.e. $M = kt^{1/2}$.

The main formulation factors by which the release rate from a matrix system can be controlled are the amount of drug in the matrix, the porosity of the release unit, the length of the pores in the release unit (dependent on the size of the release unit and the pore tortuosity) and the solubility of the drug (which regulates the concentration gradient). The characteristics of the pore system can be affected by, for example, the addition of soluble excipients and by the compaction pressure during tabletting.

Matrix systems are traditionally designed as single-unit systems, normally tablets, prepared by tabletting. However, alternative preparation procedures are also used, especially for release units that are smaller than tablets. Examples of such techniques are extrusion, spray-congealing and casting.

Dissolution-controlled release systems

In dissolution-controlled extended-release systems the rate of dissolution in the gastrointestinal juices of the drug or another ingredient is the release-controlling process. It is obvious that a sparingly water-soluble drug can form a preparation of a dissolution-controlled extended-release type. A reduced drug solubility can be accomplished by preparing poorly soluble salts or derivatives of the

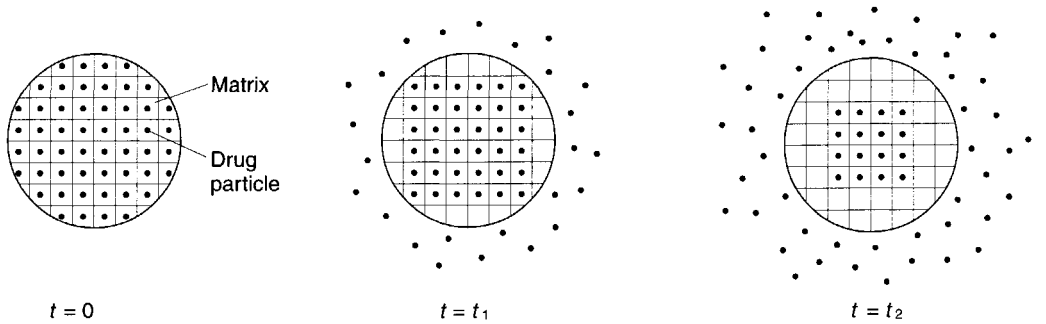


Fig. 27.14 Schematic illustration of the mechanism of drug release from a diffusion-based matrix tablet (t = time).

drug. In practice, this approach is a less common way of formulating an extended-release preparation. An alternative means to achieve extended release based on dissolution is to incorporate the drug in a slowly dissolving carrier.

Dissolution-controlled extended-release systems can also be obtained by covering drug particles with a slowly dissolving coating. The release of the drug from such units occurs in two steps:

1. The liquid that surrounds the release unit dissolves the coating (rate-limiting dissolution step).
2. The solid drug is exposed to the liquid and subsequently dissolves.

In order to obtain an extended release based on dissolution of a coating, the tablet is designed to release the drug in a series of pulses. Although this type of release is not continuous it is normally referred to as extended release, as a similar bioavailability as with continuous-release systems can often be achieved. A pulsatile drug release can be accomplished by dividing the drug dose into a number of smaller release units, which are coated in such a way that the dissolution time of the coatings will vary (Fig. 27.15). The release unit is often a nearly spherical granule about 1 mm in diameter. A variation in dissolution time of the coating can be accomplished by varying its thickness or its solubility. Release units with different release times will be mixed and formed into tablets. After disintegration of the tablet, the release units will deliver the drug in a sequence of pulses.

The procedure described here is also the most common means to prepare a delayed-release system, such as enteric-coated dosage forms. In this case dissolution is inhibited until the preparation reaches the higher pH of the small intestine, where the drug is released in a relatively short time.

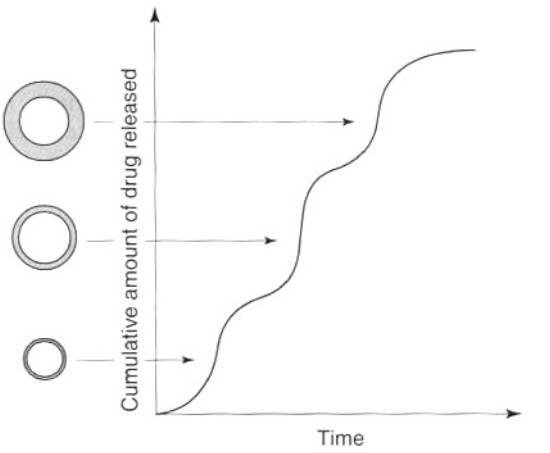


Fig. 27.15 Schematic representation of the cumulative amount of drug released from a dissolution-based (due to differences in coating thickness) pulsatile-release preparation.

Erosion-controlled release systems

In erosion-controlled extended-release systems the rate of drug release is controlled by the erosion of a matrix in which the drug is dispersed. The matrix is normally a tablet, i.e. the matrix is formed by a tabletting operation, and the system can thus be described as a single-unit system. The erosion in its simplest form can be described as a continuous liberation of matrix material (both drug and excipient) from the surface of the tablet, i.e. a surface erosion. The consequence will be a continuous reduction in tablet weight during the course of the release process (Fig. 27.16). Drug release from an erosion system can thus be described in two steps:

1. Matrix material, in which the drug is dissolved or dispersed, is liberated from the surface of the tablet.
2. The drug is subsequently exposed to the gastrointestinal fluids and mixed with (if the drug is dissolved in the matrix) or dissolved in (if the drug is suspended in the matrix) the fluid.

This release scheme is in practice a simplification, as erosion systems may combine different mechanisms for drug release. For example, the drug may be released both by erosion and by diffusion within the

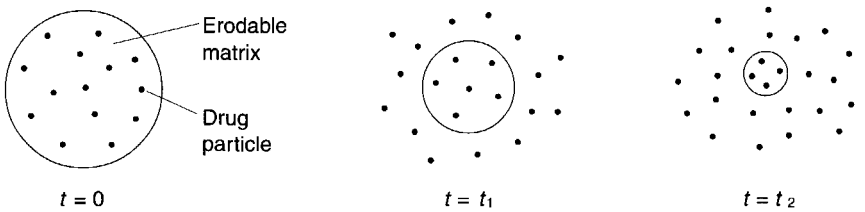


Fig. 27.16 Schematic illustration of the mechanism of drug release from an erosion tablet.

matrix. Thus, a mathematical description of drug release from an erosion system is complex. However, drug release can often approximate zero-order for a significant part of the total release time.

The eroding matrix can be formed from different substances. One example is lipids or waxes, in which the drug is dispersed. Another example is polymers that gel in contact with water (e.g. hydroxyethyl cellulose). The gel will subsequently erode and release the drug dissolved or dispersed in the gel. Diffusion of the drug in the gel may occur in parallel.

Osmosis-controlled release systems

In osmosis-controlled extended-release systems the flow of liquid into the release unit, driven by a difference in osmotic pressure between the inside and the outside of the release unit, is used as the release-controlling process. Osmosis can be defined as the flow of a solvent from a compartment with a low concentration of solute to a compartment with a high concentration. The two compartments are separated by a semipermeable membrane, which allows flow of solvent but not of the solute.

In the most simple type of osmosis-controlled drug release the following sequence of steps is involved in the release process:

1. Osmotic transport of liquid into the release unit;
2. Dissolution of drug within the release unit;
3. Convective transport of a saturated drug solution by pumping of the solution through a single orifice or through pores in the semipermeable membrane.

The pumping of the drug solution can be accomplished in different ways. One example is a tablet which includes an expansion layer, i.e. a layer of a substance that swells in contact with water, the expansion of which will press out the drug solution from the release unit. Alternatively, the increased volume of fluid inside the release unit will increase the internal pressure, and the drug solution will thus be pumped out.

If the flow rate of incoming liquid to the release unit is the rate-controlling process, the drug release rate can be described as:

$$M/t = CV/t \qquad (27.3)$$

where V is the volume of incoming liquid. The flow rate of incoming liquid under steady-state conditions is a zero-order process, and the release rate of the drug will therefore also be a zero-order process. The water flow is not affected by the flow and pH of the dissolution medium. However, the water flow rate

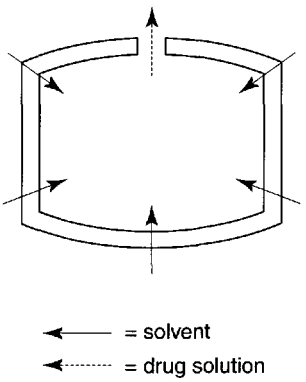


Fig. 27.17 Schematic illustration of the mechanism of drug release from an osmosis-controlled release system designed as a single-unit tablet with a single release orifice.

and hence drug release rate can be affected by a number of formulation factors, such as the osmotic pressure of the drug solution within the release unit, the drug solubility and the permeability and mechanical properties of the membrane.

Osmosis-controlled release systems can be designed as single-unit or multiple-unit tablets. In the first case the drug solution can be forced out from the tablet through a single orifice (Fig. 27.17) formed in the membrane by boring with a laser beam. Alternatively, the drug solution can flow through a number of pores formed during the uptake of water. Such pores can be formed by the dissolution of water-soluble substances in the membrane, or by straining of the membrane owing to the increased internal pressure in the release unit. In the case of multiple-unit release tablets the transport occurs in formed pores.

TABLET TESTING

Uniformity of content of active ingredient

A fundamental quality attribute for all pharmaceutical preparations is the requirement for a constant dose of drug between individual tablets. In practice, small variations between individual preparations are accepted and the limits for this variation are defined as standards in pharmacopoeias. For tablets, uniformity of dose or dose variation is tested in two separate tests: uniformity of weight and uniformity of active ingredient. These either reflect indirectly or measure directly the amount of drug substance in the tablet.

The test for uniformity of weight is carried out by collecting a sample of tablets, normally 20, from a batch and determining their individual weights. The average weight of the tablets is then calculated. The sample complies with the standard if the individual weights do not deviate from the mean more than is permitted in terms of percentage.

If the drug substance forms the greater part of the tablet mass, any weight variation obviously reflects variations in the content of active ingredient. Compliance with the standard thus helps to ensure that uniformity of dosage is achieved. However, in the case of potent drugs which are administered in low doses, the excipients form the greater part of the tablet weight and the correlation between tablet weight and amount of active ingredient can be poor (Fig. 27.18). Thus, the test for weight variation must be combined with a test for variation in content of the drug substance. Nevertheless, the test for uniformity of weight is a simple way to assess variation in drug dose, which makes the test useful as a quality control procedure during tablet production.

The test for uniformity of drug content is carried out by collecting a sample of tablets, normally 10, followed by a determination of the amount of drug in each. The average drug content is calculated and the content of the individual tablets should fall within specified limits in terms of percentage deviation from the mean.

Disintegration

As discussed above, the drug release process from immediate-release tablets often includes a step at which the tablet disintegrates into smaller fragments. In order to assess this, disintegration test methods have been developed and examples are described as official standards in pharmacopoeias.

The test is carried out by agitating a given number of tablets in an aqueous medium at a defined temperature, and the time to reach the end-point of the test is recorded. The preparation complies with the test if the time to reach this end-point is below a given limit. The end-point of the test is the point at which all visible parts of the tablets have been eliminated from a set of tubes in which the tablets have been held during agitation. The tubes are closed at the lower end by a screen and the tablet fragments formed during the disintegration are eliminated from the tubes by passing the screen openings, i.e. disintegration is considered to be achieved when no tablet fragments remain on the screen (fragments of coating may remain).

A disintegration apparatus (Fig. 27.19) consists normally of six chambers, i.e. tubes open at the upper end and closed by a screen at the lower. Before disintegration testing, one tablet is placed in each tube and normally a plastic disc is placed upon it. The tubes are placed in a water bath and raised and lowered at a constant frequency in the water in such a way that at the highest position of the tubes, the screen remains below the surface of the water.

Tests for disintegration do not normally seek to establish a correlation with in vivo behaviour. Thus, compliance with the specification is no guarantee of an acceptable release and uptake of the drug in vivo and hence an acceptable clinical effect. However, it is reasonable that a preparation that fails to comply with the test is unlikely to be efficacious. Disintegration tests are, however, useful as a means to assess the potential importance of formulation and process variables on the biopharmaceutical properties of the tablet, and as

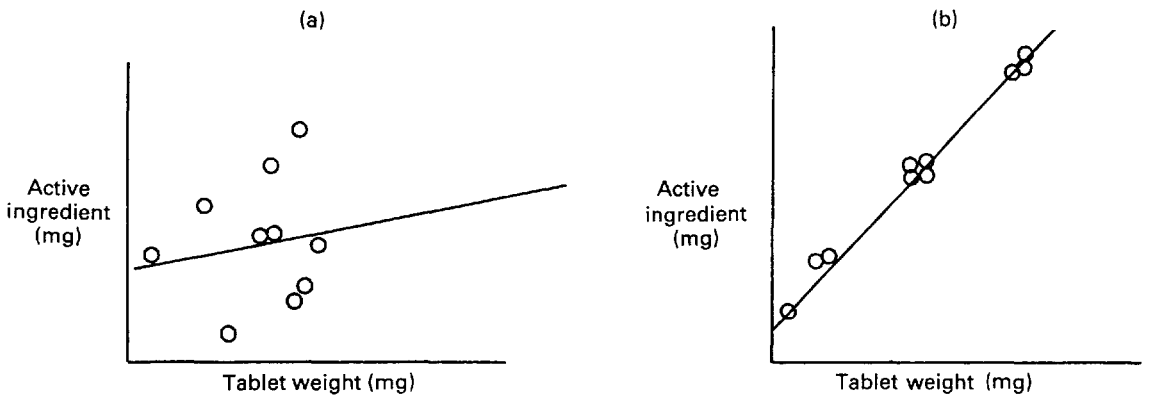


Fig. 27.18 Correlation between amount of active ingredient and tablet weight for (a) a low dose (drug content 23% of tablet weight) and (b) a high dose (drug content 90% of tablet weight) tablet. (From Airth, J.M., Bray, D.F., and Radecka, C. (1967). J. Pharm. Sci., 56, 233–235.

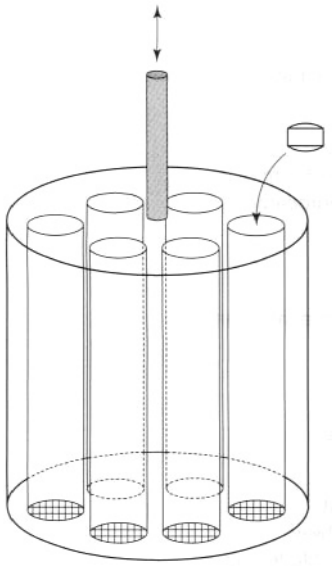


Fig. 27.19 Diagram of a disintegration instrument for the testing of tablet disintegration time.

a control procedure to evaluate the quality reproducibility of the tablet during production.

Dissolution

Dissolution testing is the most important way to study, under in vitro conditions, the release of a drug from a solid dosage form, and thus represents an important tool to assess factors that affect the bioavailability of a drug from a solid preparation. During a dissolution test the cumulative amount of drug that passes into solution is studied as a function of time. The test thus describes the overall rate of all the processes involved in the release of the drug into a bioavailable form.

Dissolution studies are carried out for several reasons:

- To evaluate the potential effect of formulation and process variables on the bioavailability of a drug;
- To ensure that preparations comply with product specifications;
- To indicate the performance of the preparation under in vivo conditions.

This last point requires that in vitro dissolution data correlate with the in vivo performance of the dosage form, which must be experimentally verified. The term in vitro/in vivo correlation in this context is related to the correlation between in vitro dissolution and the release or uptake of the drug in vivo. The establishment of such a correlation is one of the most important aspects of a dissolution test for a preparation under formulation development, and is discussed further in Chapter 18.

Dissolution is accomplished by locating the tablet in a chamber containing a flowing dissolution medium. So that the method is reproducible, all factors that can affect the dissolution process must be standardized. This includes factors that affect the solubility of the substance (i.e. the composition and temperature of the dissolution medium) and others that affect the dissolution process (such as the concentration of dissolved substance in, and the flow conditions of, the fluid in the dissolution chamber). Normally, the concentration of the drug substance in the bulk of the dissolution medium shall not exceed 10% of the solubility of the drug, i.e. sink conditions. Under sink conditions, the concentration gradient between the diffusion layer surrounding the solid phase and the concentration in the bulk of the dissolution medium is often assumed to be constant.

A number of official and unofficial methods exist for dissolution testing, which can be applied to both drug substances and formulated preparations. With respect to preparations, the main test methods are based on forced convection of the dissolution medium and can be classified into two groups: stirred-vessel methods and continuous-flow methods.

Stirred-vessel methods

The most important stirred-vessel methods are the paddle method (Fig. 27.20) and the rotating-basket method (Fig. 27.21). Details of these can be found in official monographs in the European or US Pharmacopoeias. Both use the same type of vessel, which is filled with a dissolution medium of controlled volume and temperature. In the paddle method, the tablet is placed in the vessel and the dissolution medium is agitated by a rotating paddle. In the rotating-basket method, the tablet is placed in a small basket formed from a screen. This is then immersed in the dissolution medium and rotated at a given speed.

Continuous-flow methods

In the continuous-flow method the preparation is held within a flow cell, through which the dissolution medium is pumped at a controlled rate from a large reservoir. The liquid which has passed the flow cell is collected for analysis of drug content. The continuous-flow cell method may have advantages over stirred-vessel methods, e.g. it maintains sink conditions throughout the experiment and avoids floating of the preparation.

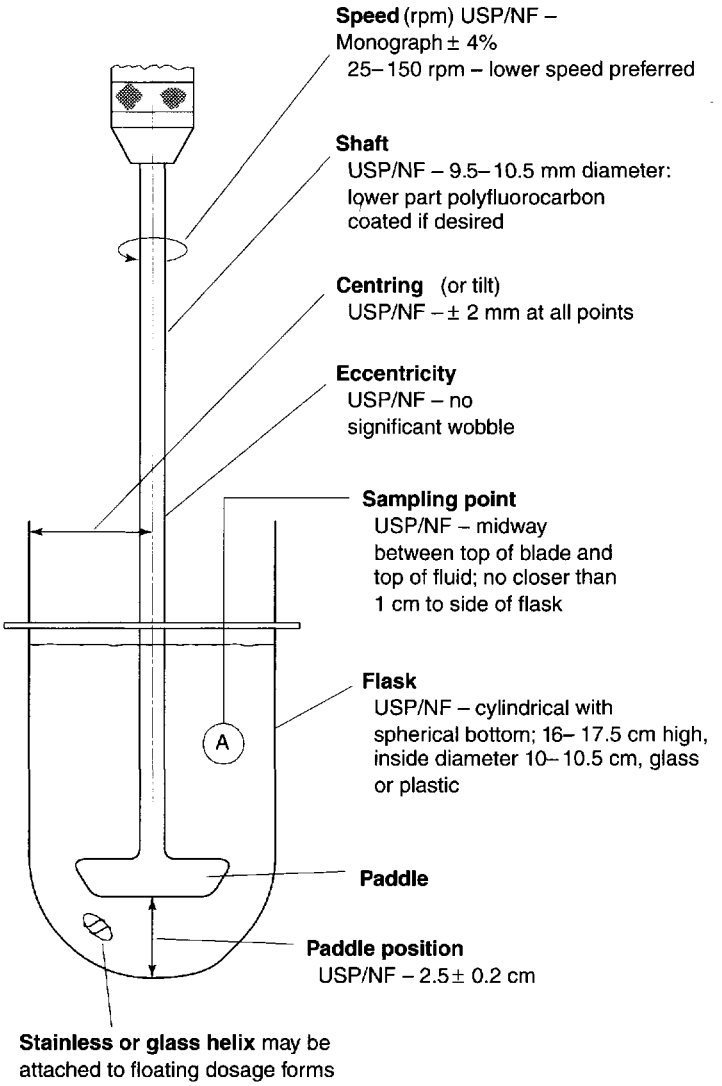


Fig. 27.20 Diagram of a dissolution instrument based on the rotating paddle method for the testing of tablet dissolution rate. (From Banakar, U.V. Pharmaceutical Dissolution Testing. Marcel Dekker, Inc., New York 1992.)

The amount of drug dissolved is normally analysed more or less continuously as the concentration in the vessel at a series of consecutive times. However, sometimes a single measurement can be performed if required in the Pharmacopoeia or product specification, i.e. the amount of drug dissolved within a certain time period is determined.

The composition of the dissolution medium might vary between different test situations. Pure water may be used, but in many cases a medium that shows a closer resemblance to some physiological fluid is used. In such media the pH and ionic strength can be controlled, and surface-active agents might be added to affect the surface tension of the liquid and the solubility of the drug. Such fluids are often referred to as simulated gastric or intestinal fluids. Also, other dissolution media might be used, such as solvent mixtures, if the solubility of the drug is very low.

Mechanical strength

The mechanical strength of a tablet is associated with the resistance of the solid specimen towards fracturing and attrition. An acceptable tablet must remain intact during handling between production and administration. Thus, an integrated part of the

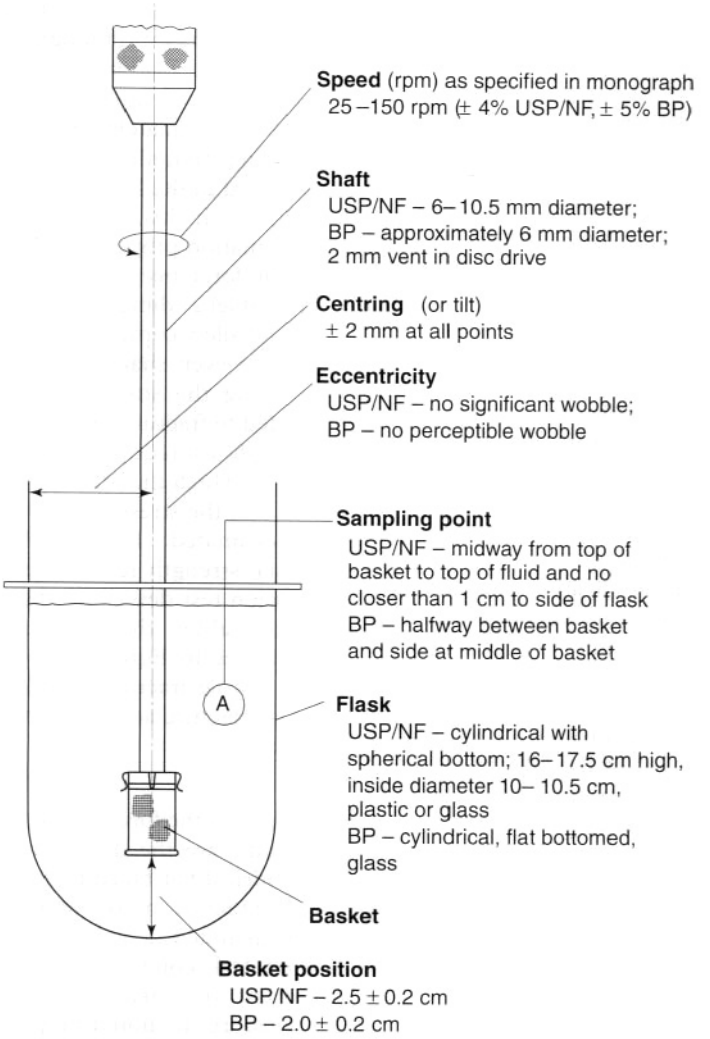


Fig. 27.21 Diagram of a dissolution instrument based on the rotating-basket method for the testing of tablet dissolution rate. (From Banakar, U.V. Pharmaceutical Dissolution Testing. Marcel Dekker, Inc., New York 1992.)

formulation and production of tablets is the determination of their mechanical strength. Such testing is carried out for several reasons, such as:

- To assess the importance of formulation and production variables for the resistance of a tablet towards fracturing and attrition during formulation work, process design and scaling up;
- To control the quality of tablets during production (in-process and batch control);
- To characterize the fundamental mechanical properties of materials used in tablet formulation.

A number of methods are available for measuring mechanical strength and they give different results. Especially for the use of strength data to assess material properties, a number of test methods originating from materials science are used, such as beam bending and uniaxial tensile testing. In this context also the hardness of a tablet can be measured by indentation. The hardness of a specimen is associated with its resistance to local permanent deformation, and is thus not a measure of the resistance of the tablet towards fracturing.

The most commonly used methods for strength testing can be subcategorized into two main groups: attrition-resistance methods and fracture-resistance methods.

Attrition-resistance methods

The idea behind attrition resistance methods is to mimic the kind of forces to which a tablet is subjected during handling between its production and its administration. These are also referred to as friability tests: a friable tablet is one that is prone to erode mechanically during handling. During handling, tablets are subjected to stresses from collisions and tablets sliding towards one another and other solid surfaces, which can result in the removal of small fragments and particles from the tablet surface. The result will be a progressive reduction in tablet weight and a change in its appearance. Such attrition can occur even though the stresses are not high enough to break or fracture the tablet into smaller pieces. Thus, an important property of a tablet is its ability to resist attrition so as to ensure that the correct amount of drug is administered and that the appearance of the tablet does not change during handling. Another application of a friability method is to detect incipient capping, as tablets with no visible defects can cap or laminate when stressed by an attrition method, e.g. a rotating cylinder.

The most common experimental procedure to determine attrition resistance involves the rotation of tablets in a cylinder followed by the determination of weight loss after a given number of rotations. Another approach is to shake tablets intensively in a jar of similar dimensions to a pack-jar. Normally, weight loss of less than 1% during a friability test is required. In addition, the tablets should not show capping or cracking during such testing.

Fracture-resistance methods

Analysis of the fracture resistance of tablets involves the application of a load on the tablet and the determination of the force needed to fracture or break the specimen along its diameter. In order to obtain a controlled loading, care must be taken to ensure that the load is applied under defined and reproducible conditions in terms of the type of load applied (compression, pulling, twisting etc.) and the loading rate.

For compressive loading of tablets, the test is simple and reproducible under controlled conditions, and the diametric compression test has therefore a broad use during formulation development and tablet production. In such compression testing the tablet is placed against a platen and the load is applied along its diameter by a movable platen. The tablet fails ideally along its diameter, i.e. parallel to the compression load, in a single fracture into two pieces of similar size (Fig. 27.22), and the fracture force is recorded. This mode of failure is actually a tensile failure even though it is accomplished here by compressive loading. The force needed to fracture the tablet by diametral compression is often somewhat unfortunately referred to as the crushing or breaking strength of the tablet. The term hardness is also used in the literature to denote the failure force, which is in this context incorrect as hardness is a deformation property of a solid.

The force needed to fracture a tablet depends on the tablet's dimensions. An ideal test, however, should allow comparison between tablets of different sizes or even shapes. This can be accomplished by assessing the strength of the tablet, i.e. the force needed to fracture the tablet per unit fracture area. A strength test requires that the fracture mode (i.e. the way by which the crack is formed) can be controlled and that the stress state along the fracture plane can be estimated. The simplest and most common tensile strength test is the indirect diametral compression test described above. For a cylindrical flat-faced tablet the tensile strength (σ_t) can be calculated by Eqn 27.4, provided that the tablet fails in a tensile fracture mode characterized by a single linear fracture across the diameter of the cylindrical specimen:

$$\sigma_t = 2F/\pi D\, t \qquad (27.4)$$

where F is the force needed to fracture the tablet and D and t are the diameter and the thickness of the cylindrical flat-faced tablet, respectively.

In practice, more complicated failure characteristics than tensile failure are often obtained during diametral compression (Fig. 27.23), which will prevent the strict application of the calculation procedure. It should be pointed out that the tensile strength of convex-faced tablets can also be calculated by using other equations.

An alternative procedure to measure the tensile strength of a tablet is to directly pull the tablet apart

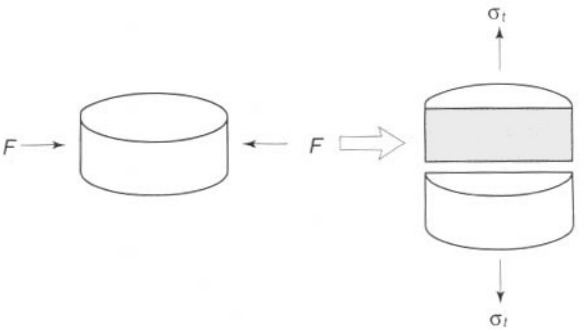


Fig. 27.22 Illustration of the tensile failure of a tablet during diametral compression.

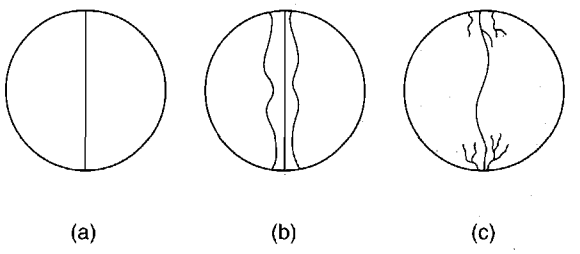


Fig. 27.23 Examples of different types of failure induced by diametral compression. (a) Simple tensile failure. (b) Triple cleft failure. (c) Failure due to shear at platen edges. (From Davies, P.N. and Newton, J.M. In: Pharmaceutical Powder Compaction Technology (Eds. Alderborn G. and Nyström, C.), Marcel Dekker Inc., New York 1996.)

by the application of stresses along its main axes until fracture occurs, i.e. a direct axial tensile test. The use of this method is primarily to detect weaknesses in the compact in the axial direction, which is an indication of capping or lamination tendencies in the tablet. Thus, the strength value obtained by this procedure indicates weak zones in the tablet rather than the mean strength of the whole tablet.

FUNDAMENTAL ASPECTS OF THE COMPRESSION OF POWDERS

Mechanisms of compression of particles

The compressibility of a powder is defined as its propensity, when held within a confined space, to reduce in volume while loaded. The compression of a powder bed is normally described as a sequence of processes. Initially, the particles in the die are rearranged, resulting in a closer packing structure and reduced porosity. At a certain load the reduced space and the increased interparticulate friction will prevent any further interparticulate movement. The subsequent reduction of the tablet volume is therefore associated with changes in the dimensions of the particles.

Particles, either whole or a part, can change their shape temporarily by elastic deformation and permanently by plastic deformation (Fig. 27.24). Particles can also fracture into a number of smaller, discrete particles, i.e. particle fragmentation. The particle fragments can then find new positions, which will further decrease the volume of the powder bed. When the applied pressure is further increased the smaller particles formed could again undergo deformation. Thus, one single particle may undergo this cycle of events several times during one compression. As a

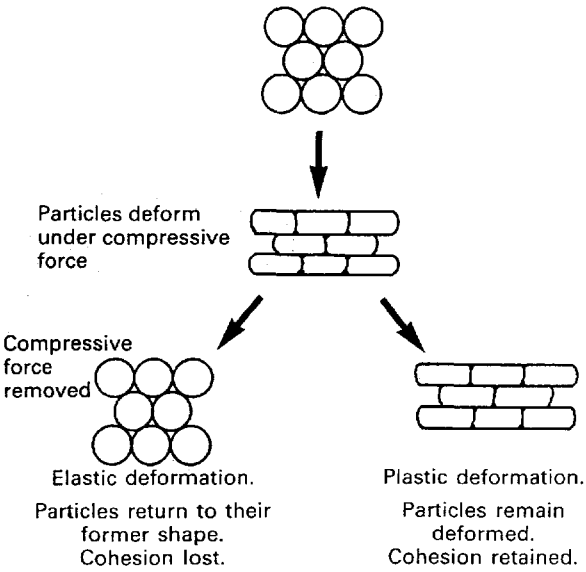


Fig. 27.24 Schematic illustration of particle deformation, elastic and plastic, during compression. (From Armstrong, N.A. Mfg. Chem. October, 64 1982.)

consequence of compression particle surfaces are brought into close proximity to each other and particle–particle bonds can be formed.

Elastic and plastic deformation of particles are a time-independent processes, i.e. the degree of deformation is related to the applied stress and not the time of loading. However, it can also be time-dependent, i.e. the degree of deformation is related to the applied stress *and* the time of loading. This deformation behaviour is referred to as viscoelastic and viscous deformation of a material. The consequence is that the compression behaviour of a material might depend on the loading conditions during the formation of a tablet in terms of the punch displacement–time relationship. Many pharmaceutical substances seem to have a viscous character, i.e. be strain-rate sensitive, and the properties of the tablet is thus dependent on the punch displacement–time relationship for the compression process.

Elastic deformation can be described as a densification of the particle due to a small movement of the cluster of molecules or ions that forms the particle, e.g. a crystal lattice or a cluster of disordered molecules. Plastic deformation is considered to occur by the sliding of molecules along slip planes within the particles. For real crystals, such slip planes are formed at defects in the crystal lattice, especially dislocations.

The majority of powders handled in pharmaceutical production consist not of non-porous primary

particles but rather of granules, i.e. porous secondary particles formed from small dense primary particles. For granules, a larger number of processes are involved in their compression. These can be classified into two groups:

- Physical changes in the granules, i.e. the secondary particles;
- Physical changes in the primary particles from which the granules are formed.

The latter concern changes in the dimensions of the primary particles due to elastic and plastic deformation and fragmentation. Such processes may be significant for the strength of tablets. It is, for example, common for a capping-prone substance, when compacted as dense particles, to also be prone to cap or laminate during compaction in the form of granules, such as substrate-binder granules. However, in terms of the evolution of the tablet structure, the physical changes in the granules that occur during compression are of primary importance.

At low compression forces the reduction in volume of the bed of granules can occur by a rearrangement within the die. However, granules are normally fairly coarse, which means that they spontaneously form a powder bed of relatively low voidage (i.e. the porosity of the intergranular spaces). Therefore, this initial rearrangement phase is probably of limited importance with respect to the total change in bed volume of the mass. With increased loading, a further reduction in bed volume therefore requires changes in the structure of the granules. The granules can deform, both elastically and permanently, but also densify, i.e. reduce their intragranular porosity. By these processes granules can still be described as coherent units, but their shape and porosity will change.

Granules can also be broken down into smaller units by different mechanisms:

1. Primary particles might be removed from the surface of granules when they slide against each other or against the die wall. This can be described as erosion or attrition, rather than fracturing. This mechanism occurs primarily for granules with a rough surface texture.
2. Granules can fracture into a number of smaller ones i.e. granule fragmentation.

Studies on the compression properties of granules formed from pharmaceutical substances have indicated that granules are not prone to fracture into smaller units during compression over a normal range of applied pressures. Thus, permanent deformation and densification dominate the compression event. However, for irregular, rough granules some attrition might occur. Deformation and densification of granules have been suggested to occur by the repositioning of primary particles within the granules, i.e. these processes involves an internal flow of primary particles. In this context, the terms ***degree*** and ***mode*** of deformation have been used to describe granule deformation. Degree of deformation refers to some quantitative change in the shape of the granules, whereas mode of deformation refers to the type of shape change, such as a flattening of the granule, or a more complicated shape change towards irregular granules.

The dominating compression mechanisms for dense particles and granules are summarized in Table 27.2. The relative occurrence of fragmentation and deformation of solid particles during compression is related to the fundamental mechanical characteristics of the substance, such as their elasticity and plasticity. For granules, both the mechanical properties of the primary particles from which the granules are formed and the physical structure of the granule, such as their porosity and shape, will affect the relative occurrence of each compression mechanism.

Evaluation of compression behaviour

Procedures

The procedures used in research and development work to evaluate the compression behaviour of particles and the mechanisms of compression involved in the volume reduction process are of two types:

- Characterization of ejected tablets;
- Characterization of the compression and decompression events.

Concerning the characterization of ejected tablets, the most important procedures used are inspection and the determination of the pore structure of the

Table 27.2 Dominating compression mechanisms for dense particles and granules (porous particles)

Dense particles	Granules
Repositioning of particles	Repositioning of granules
Particle deformation	Granule deformation
elastic	(permanent)
plastic	Granule densification
viscous/viscoelastic	Granule attrition
Particle fragmentation	Deformation of primary particles

tablet, in terms of a mean pore size, pore size distribution and specific surface area. A less common approach is to calculate ratios between the mechanical strengths of tablets measured in different directions.

Concerning characterization of the compression and decompression events, these procedures are based on relationships between parameters that can be derived from the compaction process (Table 27.3). Some of the most common approaches used in this context are described below.

Inspection of tablets

The inspection of tablets, e.g. by scanning electron microscopy, is an important means to study changes in the physical properties of particles during compression. Such changes include fragmentation into smaller particles, permanent shape changes due to deformation, and finally, the formation of cracks within the particles. Such inspection will also give information about the relative positions of particles within the tablet and hence the interparticulate pore structure. The fracture path during strength testing, i.e. failure around or across the particles, can also be estimated from inspection of the tablet fracture surface.

In addition to the inspection of intact tablets, studies of the fragmentation of particles during compression can be obtained by analysing the size and size distribution of particles obtained by deaggregation of a tablet. Such deaggregation can occur spontaneously by disintegration of the tablet in a liquid, or be created mechanically. Studies on such deaggregated tablets have indicated that powder compression can effectively reduce the size of particles and result in a wider particle size distribution of the cohered particles within a tablet.

Table 27.3 Parameters used in procedures to describe compression and decompression events

Upper punch force/pressure versus compression time*
Lower punch force/pressure versus compression time†
Upper punch force/pressure versus lower punch force/pressure
Upper punch force versus die-wall force
Punch force versus punch displacement (mainly upper punch)
Tablet volume versus upper punch pressure/force
Tablet porosity versus upper punch pressure/force

* Used both during ordinary compression and also as prolonged loading after maximum applied force/pressure has been reached (referred to as stress relaxation measurements).
† Used primarily to describe the ejection phase.

Pore structure and specific surface area of tablets

One of the most important ways to study the evolution of tablet structure during compression is to measure some characteristic of the pore structure of the tablet. Information on pore size distribution can be obtained by mercury intrusion measurements and by gas adsorption–desorption. However, the most common way to evaluate the pore system of a tablet has been to measure the surface area of the tablet by air permeability or gas adsorption. The former has also been used to derive an indication of the mean pore size in a tablet.

Fragmentation is a size reduction process and so can be assessed by measuring the specific surface area of a particulate solid before and after compaction, or measuring changes in tablet surface area with compaction pressure. The surface area of tablets can be determined by several procedures, including gas adsorption and mercury intrusion. Both these methods can provide useful information on particle fragmentation. However, by these methods it is difficult to differentiate between inter- and intraparticulate pores, and the procedure might thus not reflect particle fragmentation only, but also the formation or closure of cracks or intraparticulate pores.

A method which measures the external surface area is thus advantageous in this context. Air permeametry is one such method which also has the advantage of being simple and fast. The procedure involves the formation of tablets at a series of applied pressures, followed by the measurement of the tablet's specific surface area. The slope of the relationship between tablet surface area and applied pressure represents the degree of fragmentation and can be used to classify materials with respect to their fragmentation propensity (Fig. 27.25). It should be pointed out that the calculation of tablet surface area from air permeability measurements may give erroneous values as a result of the assumptions made in the derivation of the calculation procedure. In spite of this, the method has been shown to give useful data in terms of describing the fragmentation propensity of a substance.

The relationship between tablet surface area and applied pressure is, however, strongly dependent on the original surface area of the powder, i.e. the tablet

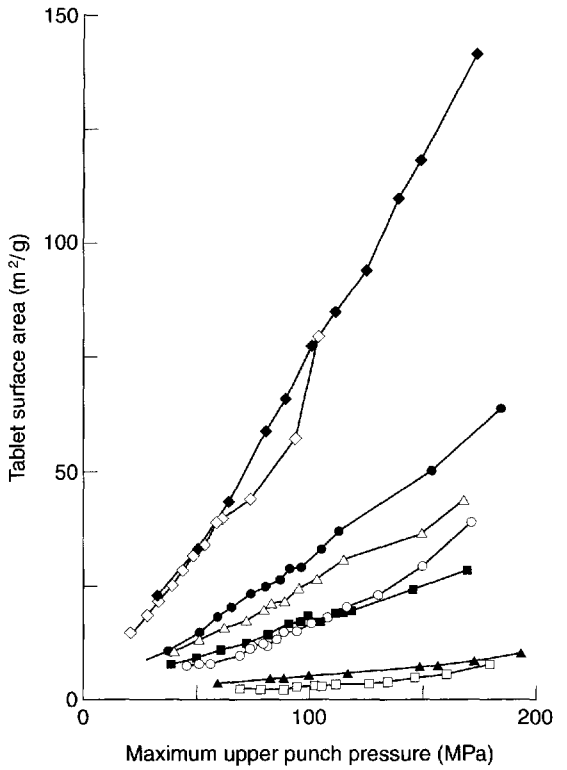


Fig. 27.25 The tablet surface area, measured by air permeametry, as a function of compaction pressure for a series of pharmaceutical substances. (From Alderborn, G., Pasanen, K., and Nyström, C. Int. J. Pharm., 23, 79 1985. □ sodium chloride, ▲ sodium bicarbonate, ○ saccharose, ■ sodium citrate, △ ascorbic acid, ● lactose, ◇ paracetamol, ◆ Emcompress.

surface area increases more markedly with applied pressure when the original particle size was smaller. Attempts have been made in the literature to derive an expression similar to those describing the size reduction of particles during milling (see Chapter 11), by which a measure of the propensity of particles to fragment independent of the original powder surface area can be calculated.

It is generally assumed that a change in the size of a particle affects the mechanics of particle deformation, i.e. how a particle responds to an applied load. Such a size-related change in the mechanics of particles can, for example, be attributed to a reduced probability of the presence of flaws in the crystal structure at which a catastrophic failure is initiated. It seems possible, therefore, that at a limiting particle size fragmentation might cease. Examples of such transitions from a brittle to a plastic behaviour have been reported, and the particle size at which this transition takes place is referred to as the critical particle size. An example is that of α-lactose monohydrate crystals, where this transition takes place at about 20 μm. This critical size has been suggested to vary markedly between different substances.

The use of surface area data to estimate particle fragmentation from gas adsorption and permeability measurements can only be applied to powders consisting of solid particles, and not to those consisting of porous particles, such as granules. The problem for the latter type is related to the internal surface area and the intraparticulate porosity of the granules.

As granules are porous particles, the pore structure of a bed of granules is dualistic, i.e. there will be pores both within and between the granules. The compression process will dramatically affect the pore structure, and for tablets prepared from granules it is in practice difficult to distinguish between inter- and intragranular pores. It seems, though, that granules tend to keep their integrity during compaction, and hence the pore structure of the tablet is principally dualistic. It has been shown that the porosity of the intragranular pore space constitutes a significant portion of the total porosity of a tablet formed from granules. As granules densify during compression, measures of their porosity before compression is not sufficient compensation for the intragranular part of the pore system.

In order to avoid the problem of defining the porosity of the intergranular pore space, a simplified approach based on air permeability measurements has been used instead to characterize the compression behaviour of granules. In this case, a tablet permeability coefficient is measured and studied in relation to the applied pressure. By this procedure, the change in intergranular tablet pore structure with applied pressure can be assessed, which is suggested to reflect the deformation and densification behaviour of granules during tabletting.

Force–displacement profiles

The relationship between upper punch force and upper punch displacement during compression, often referred to as force–displacement profile, has been used as a means to derive information on the compression behaviour of a powder and to make predictions on its tablet-forming ability. The area under a force–displacement curve represents the work or energy involved in the compression process. Different procedures have been used to analyse the curves.

One suggested approach is based on the division of the force–displacement curve into different regions (denoted E1, E2 and E3 in Fig. 27.26). It has been suggested that the areas of E1 and E3 should be as small as possible if the powder will perform well in a tabletting operation and give tablets of a high

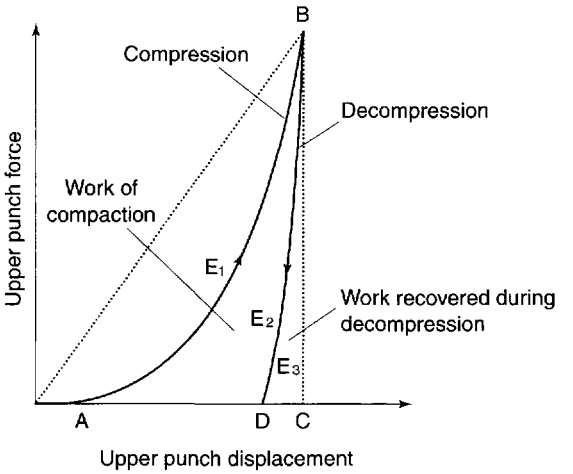


Fig. 27.26 The relationship between upper punch force and upper punch displacement during compression and decompression of a powder. (From Ragnarsson, G. In: Pharmaceutical Powder Compaction Technology (Eds. Alderborn G. and Nyström, C.), Marcel Dekker Inc., New York 1996.)

mechanical strength. An alternative proposed approach is based on mathematical analysis of the force–displacement curve from the compression phase, e.g. in terms of a hyperbolic function.

Force–displacement curves have some use in pharmaceutical development as an indicator of the tablet-forming ability of powders, including the assessment of the elastic properties of materials from the decompression curve. It can also be used as a means to monitor the compression behaviour of a substance in order to document and evaluate reproducibility between batches. However, the interpretation of the force–displacement relationship in terms of mechanisms of particle compression, or compression mechanics, is not clarified, which limits the use of force–displacement curves in fundamental compression studies.

Force–displacement measurements have also been used in fundamental studies on the energy conditions during compaction of powders, i.e. a thermodynamic analysis of the process of compact formation. The energy applied to the powder can be calculated from the area under the force–displacement curve. This compaction energy is used to overcome friction between particles, to deform particles both permanently and reversibly, and to create new particle surfaces by fragmentation. The thermal energy released during compaction can be assessed by calorimetry, i.e. the die is constructed as a calorimeter. The heat released during compression is the result of particle deformation – i.e. energy is consumed during deformation and thereafter partly released when the deformation is completed – and the result of the formation of interparticulate bonds.

Data have been reported indicating that the net effect of a compaction process is exothermal, i.e. more thermal energy is released during compaction than is applied to the powder in terms of mechanical energy. The main explanation for this is released bonding energy in the form of heat due to the formation of bonds between particles.

Tablet volume–applied pressure profiles

In both engineering and pharmaceutical sciences, the relationship between volume and applied pressure during compression is the main approach to deriving a mathematical representation of the compression process. A large number of tablet volume–applied pressure relationships exist. In addition to tablet volume and applied pressure parameters, such expressions include some constants which often are defined in physical terms. However, only for a few equations has the physical significance of the constants been generally accepted. Among these, the most recognized expression in both engineering and pharmaceutical science is the tablet porosity–applied pressure function according to Heckel.

Heckel equation Tablet porosity can be measured either on an ejected tablet or on a powder column under load, i.e. in die. The latter approach is more common as it can be performed rapidly with a limited amount of powder. A problem might be that the compression time is different at each pressure, which could affect the profile for materials having pronounced time-dependent compression behaviour.

The compression of a powder can be described in terms of a first-order reaction where the pores are the reactant and the densification the product. Based on this assumption, the following expression was derived:

$$\ln (1/E) = KP + A$$

where E is the tablet porosity, P the applied pressure, A a constant suggested to reflect particle rearrangement and fragmentation, and K the slope of the linear part of the relationship which is suggested to reflect the deformation of particles during compression. The reciprocal of the slope value K is often calculated and considered to represent the ***yield stress*** or ***yield pressure*** (P_y) for the particles, i.e.:

$$\ln (1/E) = (P/P_y) + A$$

The yield stress is defined as the stress at which particle plastic deformation is initiated. To be able to use the Heckel yield pressure parameter to compare

different substances, it is important to standardize the experimental conditions, such as tablet dimensions and speed of compaction.

Figure 27.27 shows a typical Heckel profile. This often shows an initial curvature (phase I) which has been suggested to reflect particle fragmentation and repositioning. Thereafter, the relationship is often linear over a substantial range of applied pressures (phase II), and thus obeys the expression. From the gradient of this linear part the yield pressure can be calculated, which is thus a measure of the particle plasticity. Finally, during decompression an expansion in tablet height is represented by increased tablet porosity (phase III). From this decompression phase a measure of the particle elasticity can be calculated as the relative change in tablet porosity or height.

Strain-rate sensitivity Another proposed use of yield pressure values from Heckel profiles is to assess the time-dependent deformation properties of particles during compression by comparing yield pressure values derived under compression at different punch velocities. A term denoted the ***strain-rate sensitivity*** (SRS) has been proposed (Roberts and Rowe 1985) as a characteristic of the time dependency of a powder:

$$\text{SRS} = (P_y' - P_y'')/P_y''$$

where P_y' is the yield pressure derived at a high punch velocity and P_y'' is that derived at a low punch velocity. P_y' is normally higher than P_y'' and the SRS is thus a positive value.

The discussion on the use of Heckel profiles to derive a measure of the compression yield pressure is applicable to the compression of powders consisting of solid particles. It should be emphasized that the interpretation of $1/K$ in terms of a mean yield stress for the particles is under debate. Nevertheless, support has been presented that such an interpretation is valid for solid (non-porous) particles. For porous particles, i.e. granules and pellets, the Heckel procedure is inadequate for the derivation of a measure of deformability or granule strength. The problem of applying the Heckel approach to the compression of porous particles is related to the need to assess the porosity of the reactant pore system. The pore space of interest in relation to the Heckel equation is intergranular, and the problem of quantifying this is discussed above.

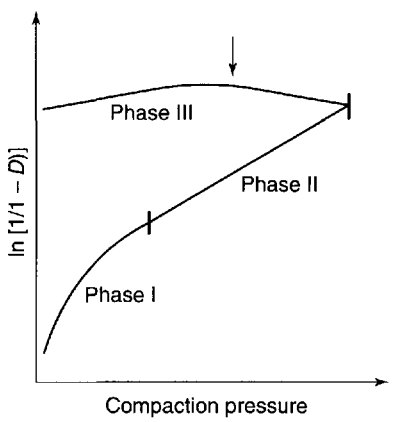


Fig. 27.27 A typical example of a Heckel profile during compression and decompression of a powder. (From Duberg, M. and Nyström, C. Powder Technol. 46, 67, 1986.)

Kawakita equation A promising means of assessing the compression mechanics of granules is to calculate a compression shear strength from the ***Kawakita equation***. This was derived from the assumption that, during powder compression in a confined space, the system is in equilibrium at all stages, so that the product of a pressure term and a volume term is constant. The equation can be written in the following linear form:

$$P/C = (1/ab) + (P/a)$$

where P is applied pressure, C the degree of volume reduction and a and b are constants. The degree of volume reduction relates the initial height of the powder column (h_o) to the height of the powder column (the compact) at an applied pressure P (h_p) as follows:

$$C = (h_o - h_p)/h_o$$

The equation has been applied primarily to powders of solid particles. However, it has been suggested (Adams et al 1994) that the compression parameter $1/b$ corresponds to the strength of granules in terms of a compression strength. The procedure thus represents a possible means to characterize the mechanical property of granules from a compression experiment.

Evaluation of die-wall friction during compression

Friction is a serious problem during tabletting. A series of procedures has thus been developed with the aim of assessing the friction between the powder or tablet and the die wall during compression and ejection, which can be used during tablet formulation to evaluate lubricants. These methods are based mainly on the use of force signals during powder compression or tablet ejection. The most common type of compression situation used in this context is

to use a single-punch press with a movable upper punch and a stationary lower punch. In such a rig the force is applied by the upper punch and transmitted axially to the lower punch, and also laterally to the die. The ejection of the tablet involves the application of an ejection force by the lower punch. Typical force profiles during compression in a single-punch press with a stationary lower punch are given in Figure 27.28.

When the descending upper punch establishes contact with the powder bed in the die, the force increases with compression time. The applied force rises to a maximum value and thereafter decreases during the decompression phase to zero. Parallel with the force trace from the upper punch, force traces from the lower punch and the die will be obtained. These can be described as transmitted forces and the force values are thus generally lower than the applied force. The force transmitted from the upper punch to the lower is considered to depend on a number of factors, including the friction between the powder and the die wall. These factors can be summarized in the following expression:

$$F_a = F_b\, e(KL/D)$$

where F_a and F_b are applied and transmitted forces, L and D are the length and diameter of the powder column within the cylindrical die (Fig. 27.29) and K is a constant. The constant K is a function of the friction coefficient between particles and the die wall. Thus, the transmission of force from the upper to the lower punch depends on the friction between the powder and the die wall. Both the difference in transmitted force, i.e. upper punch force–lower punch force, and the ratio between the upper and lower punch force, i.e. lower punch force/upper punch force (often denoted the R value), are used as measures of die-wall friction during compression. For a well lubricated powder the force transmission corresponds to $R > 0.9$.

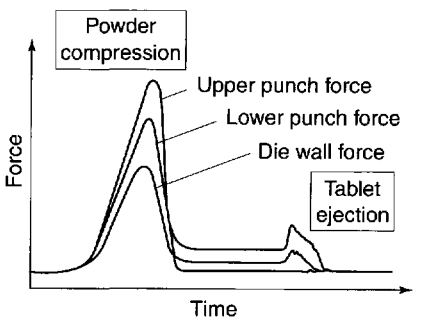


Fig. 27.28 Force–time signals (from punches and die) during uniaxial powder compression.

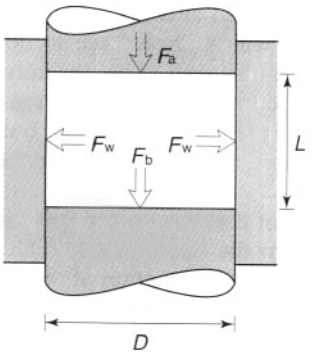


Fig. 27.29 Schematic illustration of punch and die-wall forces involved during uniaxial powder compression in a cylindrical die

In addition to studies on transmitted forces during uniaxial compression of a powder, studies have been performed with the intention of describing the distribution of the compression pressure within the powder column in a more detailed way. A complex pressure pattern will be developed during the compression, an example of which is given in Figure 27.30. This distribution in pressure will probably be associated with local variations in porosity, pore size and strength within the tablet, caused by, for example, pressure-related variations in the degree of particle deformation within the tablet.

After the upper punch has lost contact with the tablet and its force has consequently decreased to zero, the tablet will be positioned in the die in contact

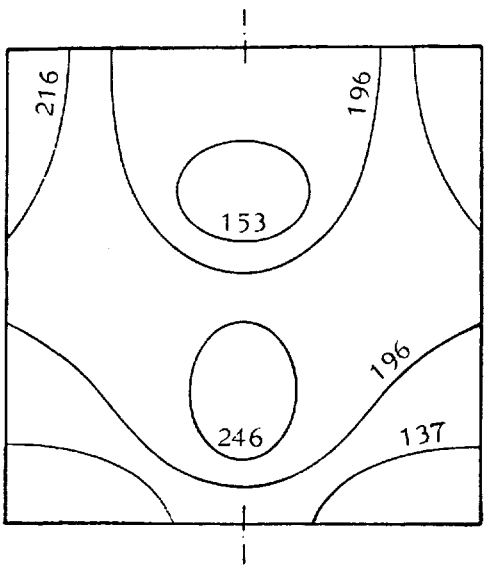


Fig. 27.30 The distribution of compression pressure (in MPa) during uniaxial powder compression (From Train, D. Trans. Inst. Chem. Engrs. 35, 258, 1957.)

with the lower punch and the die wall. In this situation, the tablet will apply a force to both the lower punch and the die wall. The magnitude of these forces is dependent on the mechanical character of the particles formed into the tablet, but also by the friction conditions at the interface between tablet and die wall.

The ejection of the tablet will result in an increased force signal from the lower punch, referred to as the ejection force. This is a function of the lateral die-wall force, but also of the friction condition at the interface between tablet and die wall. The maximum ejection force is thus also used as a measure of friction between tablet and die wall. One approach to assess friction during ejection is to calculate the dimensionless friction coefficient (μ) as the ratio between the ejection force (F_e) and the die-wall force (F_w) at the beginning of the ejection phase, i.e.:

$$\mu = F_e/F_w$$

To summarize, the following procedures are mainly used to derive measures of friction between powder or tablet and the die wall from force signals during tabletting in a single-punch press:

- Force difference between upper and lower punch;
- Force ratio between lower and upper punch;
- Maximum ejection force;
- Friction coefficient during ejection.

FUNDAMENTAL ASPECTS OF THE COMPACTION OF POWDERS

Bonding in tablets

The transformation of a powder into a tablet is fundamentally an interparticulate bonding process, i.e. the increased strength of the assembly of particles is the result of the formation of bonds between them. The nature of these bonds is traditionally subdivided into five types – known as the Rumpf classification:

1. Solid bridges
2. Bonding by liquids (capillary and surface tension forces)
3. Binder bridges (viscous binders and adsorption layers)
4. Intermolecular and electrostatic forces
5. Mechanical interlocking.

In the case of compaction of dry powders, two of the suggested types of bond are often considered to dominate the process of interparticulate bond formation, i.e. bonding due to intermolecular forces and bonding due to the formation of solid bridges. Mechanical interlocking between particles is also considered as a possible but less significant bond type in tablets.

Bonding by intermolecular forces is sometimes also known as ***adsorption bonding***, i.e. the bonds are formed when two solid surfaces are brought into intimate contact and subsequently adsorb to each other. Among the intermolecular forces, dispersion forces are considered to represent the most important bonding mechanism. This force operates in vacuum and in a gaseous or liquid environment up to a separation distance between the surfaces of approximately 10–100 nm.

The formation of solid bridges, also referred to as the ***diffusion theory of bonding***, occurs when two solids are mixed at their interface and accordingly form a continuous solid phase. Such a mixing process requires that molecules in the solid state are movable, at least temporarily, during compression. An increased molecular mobility can occur due to melting, or as a result of a glass–rubber transition of an amorphous solid phase.

Mechanical interlocking is the term used to describe a situation where strength is provided by interparticulate hooking. This phenomenon usually requires that the particles have an atypical shape, such as needle-shaped, or highly irregular and rough particles.

For tablets of a porosity in the range 5–30% it is normally assumed that bonding by adsorption is the dominant bond type between particles. In tablets formed from amorphous substances or from substances with low melting points, it is possible that solid bridges can be formed across the particle–particle interface. It is also reasonable that if tablets of a very low porosity, i.e. close to zero, are formed, particles can fuse together to a significant degree.

Often granules, i.e. secondary particles formed by the aggregation of primary particles, are handled in a tabletting operation. When granules are compacted, bonds will be formed between adjacent granule surfaces. For granules that do not include a binder the fusion of adjacent surfaces during compaction is probably not a significant bonding mechanism. Thus, intermolecular bonding forces acting between intergranular surfaces in intimate contact will probably be the dominant bond type in such tablets.

Granules often include a binder. When such binder–substrate granules are compacted it is reasonable to assume that the binder also plays an important role in the formation of intergranular bonds. The binder may fuse together locally and form binder bridges between granule surfaces which

cohere the granules to each other. Such bridges may be the result of a softening or melting of binder layers during the compression phase. These bonds can be described as solid bridges according to the Rumpf classification (see above). However, different types of adsorption bonds may be active between granule surfaces. These may be subdivided into three types: binder–binder, binder–substrate and substrate–substrate bonds.

For adsorption bonds between granules in a tablet, the location of the failure during fracturing of the tablet can vary. Fractures occurring predominantly through binder bridges between substrate particles, as well as predominantly at the interface between the binder and the substrate particle, may occur. The location of the failure has been attributed to the relative strength of the cohesive (binder bridge) and adhesive (binder-substrate interface) forces acting within the granules, which can be affected by, for example, the surface geometry of the substrate particles.

The main bond types in tablets formed from dense particles (interparticulate bonds) and from granules (intergranular bonds) are summarized in Table 27.4.

The compactability of powders and the strength of tablets

The compactability of a powder refers to its propensity to form a coherent tablet and thus represents a critical powder property in successful tabletting operations. The ability of a powder to cohere is understood in this context in a broad sense, i.e. a powder with a high compactability forms tablets with a high resistance towards fracturing and without tendencies to cap or laminate (Fig. 27.31). In practice, the most common way to assess powder compactability is to study the effect of compaction pressure on the strength of the resulting tablet, as assessed by the force needed to fracture the formed tablet while loaded diametrically, or the tensile strength of the tablet. Such relationships are often nearly linear (Fig. 27.32) above a lower pressure threshold needed to form a tablet and up to a pressure corresponding to a tablet of a few percent porosity. At low porosities the relationship between tablet strength and compaction pressure will often level out. This relationship can thus be described simply in terms of a three-region relationship characterized by lower and upper tablet strength thresholds and an intermediate region in which the tablet strength is pressure dependent in an almost linear way. However, if cracks are formed in the tablet during tabletting, e.g. during the ejection phase, this will often affect the assessed strength. Cracking and capping can often be induced at relatively high compaction pressures. This can often be reflected as a drop in the tablet strength–compaction pressure profile.

Table 27.4 Suggested predominant bond types in tablets formed from dense particles (interparticulate bonds) and from granules (intergranular bonds)

Interparticulate bonds	Intergranular bonds
Adsorption bonds (intermolecular forces)	Adsorption bonds of three types: binder–binder, binder–substrate, substrate–substrate
Solid bridges	Solid binder bridges

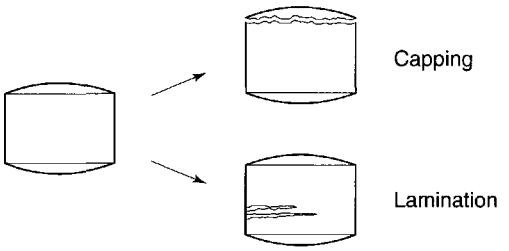


Fig. 27.31 Illustration of tablet defects referred to as capping and lamination.

By determining the tensile strength of a tablet by direct pulling in the axial direction, weaknesses in the tablet caused by small cracks can be observed as a reduction in axial tensile strength without a parallel effect on the tablet's tensile strength.

Alternatives to tablet strength–compaction pressure relationships for representing the compactability of powders are also used, such as the relationship

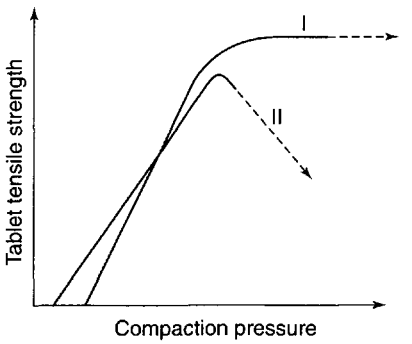


Fig. 27.32 Outline of the relationship between tablet tensile strength and compaction pressure for tablets showing no lamination (I) and for tablets showing lamination or capping (II).

between tablet strength and tablet porosity, and the relationship between tablet strength and the work done by the punches during tablet formation.

Compaction is fundamentally a bonding process, i.e. strength is provided by bonds formed at the interparticulate junctions or contact sites during the compression process. Studies on the structure of fractured tablets indicate that a tablet generally fails by the breakage of interparticulate bonds, i.e. an interparticulate fracture process. However, especially for tablets of low porosity, the tablet can also fracture by breakage of the particles that form the tablet, i.e. a combination of an inter- and an intraparticulate fracture process. In general terms it seems, though, that the interparticulate contacts in a tablet represent the preferred failure path during fracturing. This conclusion is applicable both to tablets formed from solid particles and to tablets formed from porous secondary particles (granules and pellets). Consequently, factors that affect the microstructure at the interparticulate junctions have been considered significant for the compactability of a powder.

Our understanding of the mechanical strength of a solid is based on the resistance of a solid body to fracture while loaded. It might seem reasonable that the sum of the bonding forces that cohere the molecules forming the solid will represent the strength of that solid. However, solids fail by a process of crack propagation, i.e. the fracture is initiated at a specific point within the solid and is thereafter propagated across a plane, thereby causing the solid to break. The consequence in terms of the strength of the solid is that the sum of the bonding forces acting over the fracture surface will be higher than the stress required to cause failure. It is known, for example, that for crystalline solids the theoretical strength due to the summation of intermolecular bonds is much higher than the measured strength of the solid.

In order to understand the strength of solids, the process of fracture has attracted considerable interest in different scientific areas. In this context important factors associated with the fracturing process and the strength of a specimen are the size of the flaw at which the crack is initiated and the resistance of the solid towards fracturing. The latter property can be described by the ***critical stress intensity factor***, which is an indication of the stress needed to propagate a crack. Another fracture mechanics parameter, which is related to the critical stress intensity factor, is the strain energy release rate, which is a measure of the energy that is released during crack propagation. By using the critical stress intensity factor, the tensile strength of the solid is considered to relate to the flaw size (c) and the critical stress intensity factor (K_{IC}) in the following way:

$$\sigma_t = f(K_{IC}/c^{1/2})$$

The critical stress intensity factor varies with tablet porosity. It has therefore been suggested that for compacts, such as tablets, factors such as the size of the particles within the tablet and the surface energy of the material will affect the critical stress intensity factor (Kendall 1988). These factors are also considered to control the interparticulate bond structure in a tablet.

Procedures to determine the critical stress intensity factor for a particulate solid have been described. Such a procedure involves normally the formation of a beam-shaped compact in which a notch is formed. When the compact is loaded, the fracture is initiated at the notch. The force needed to fracture the compact is determined and the critical stress intensity factor thereafter calculated. In order to assess a material characteristic, compacts of a series of porosities are formed and the series of values for the critical stress intensity factor subsequently determined is plotted as a function of the compact porosity (Fig. 27.33). The relationship is thereafter sometimes extrapolated to zero porosity, and the value thus derived is sometimes considered a fundamental material characteristic.

In addition to the fundamental studies on the strength of solids, indices and expressions have been derived within pharmaceutical science which can be described as indicators of the compactability

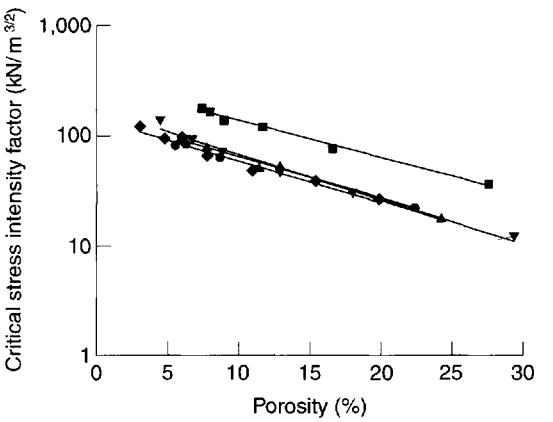


Fig. 27.33 A log-linear relationship between the critical stress intensity factor and the compact porosity for beams formed from polyethylene glycols of different molecular weight (From Al-Nasassrah, M.A., Podczeck, F., Newton, J.M. Eur. J. Pharm. Biopharm. 46, 31, 1998.)

of a powder. There are several applications of such indicators during pharmaceutical development, such as:

- the evaluation of the compactability of small amounts of particles;
- the selection of drug candidates during preformulation based on technical performance;
- the detection of batch variations of drugs and excipients;
- the selection of excipients and the evaluation of the compactability of formulations.

Examples of such indicators which have found industrial use are the indices of tabletting performance derived by Hiestand and co-workers (1996). Hiestand derived three indices of tabletting performance, among which the ***bonding index*** (BI) and the ***brittle fracture index*** (BFI) are suggested to reflect the compactability of the powder. These indices are dimensionless ratios between mechanical properties of compacts formed at some porosity. The bonding index is proposed to reflect the ability of particles to form a tablet of high tensile strength, whereas the brittle fracture index is proposed to reflect the ability of a tablet to resist fracturing and lamination during handling. These indices are defined as follows:

$$\mathrm{BI} = T/H$$

and

$$\mathrm{BFI} = (T/T_0 - 1)/2$$

where T is the tensile strength of a normal compact, T_0 is the tensile strength of a compact with a small hole and H is the hardness of the compact.

Other approaches to derive an indicator of the compactability of a powder aim to describe the microstructure of a tablet in terms of an interparticulate bond structure. They are based on the view that bond formation during compaction is significant for the development of coherence, i.e. it is postulated that the tensile strength of a tablet has some proportionality to the interparticulate bonds that act over the fracture area. The latter can be modelled in terms of, for example, the effective number of bonds and the effective contact area of the bonds. Such models can thus be described as bond summation approaches, and it is implicit that all bonds are separated simultaneously. This is not consistent with the real mode of failure of a solid and the models are thus not fundamental approaches to understanding the strength of a tablet. They can, however, be described as pragmatic models, with the aim of describing the importance of the compression behaviour of particles for the evolution of bond structure and tablet strength.

Examples of such equations have been given by Leuenberger (1982) and Alderborn and co-workers (Eriksson and Alderborn 1995, Sebhatu and Alderborn 1999). In both cases the bond structure is modelled and related to an end-point representing the maximum tensile strength (σ_t^{max}) that can be obtained for tablets of a specific powder. This maximum tensile strength can thus be described as reflecting the bond strength. Leuenberger's approach is based on the concept of effective number of interparticulate bonds in a cross-section of the tablet. It is assumed that over a cross-section of a tablet a number of bonding and non-bonding sites exist. This number depends on the applied pressure during compression (P) and the tablet relative density (ρ, which is equivalent to 1 minus the tablet porosity). In the derivation of the expression the term ***compression susceptibility*** (γ) was introduced, which described the compressibility of the powder and has the unit 1/pressure. The equation takes the following form:

$$\sigma_t = \sigma_t^{max} (1 - e [\gamma P \rho])$$

Alderborn's approach is based on the concept of effective contact area between particles in a cross-section of the tablet. In the derivation it was assumed that both particle fragmentation and particle permanent deformation are bond-forming processes, and that the former controlled the number of bonds in a tablet cross-section and the latter the area of contact between a pair of particles in that section. The expression indicates that the propensity of particles to deform irreversibly during compression is a dominant factor for the microstructure of the formed tablet, which thus is considered to be of significant importance for the resistance of a tablet towards fracturing (Fig. 27.34).

Post-compaction tablet strength changes

The compactability of a powder is normally understood in terms of the ability of particles to cohere during the compression process and hence to form a porous specimen of defined shape. However, the mechanical strength of tablets can change, increase or decrease, during storage without the application of any external mechanical force. The underlying mechanisms for such changes are often a complex function of the combination of ingredients in the tablet and the storage conditions, such as relative humidity and temperature. In order to describe

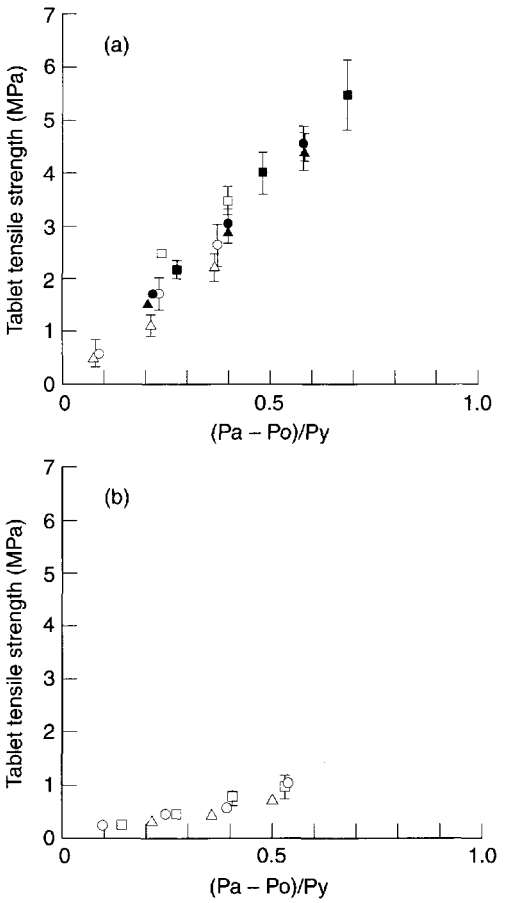


Fig. 27.34 The tensile strength of lactose tablets as a function of their microstructure, as calculated in terms of an effective contact area between the particles in the tablet by using the yield pressure from Heckel profiles (P_y). Upper graph represents tablets formed from amorphous lactose of three different particle sizes and lower graph represents tablets formed from crystalline lactose of three different particle sizes. (From Sebhatu, T. and Alderborn, G. Eur. J. Pharm. Sci. 8, 235, 1999.)

Table 27.5 Proposed mechanisms for post-compaction changes (increase or decrease) in the mechanical strength of tablets

Decreased tablet strength
Reduced bonding (intermolecular forces) due to condensation of water in tablet pores
Change in tablet microstructure due to dissolution of material in condensed water
Softening of amorphous material due to water absorption
Increased tablet strength
Formation of bonds due to crystallization of material dissolved in condensed water
Formation of bonds due to crystallization of amorphous material in rubbery state
Change in tablet microstructure due to viscous particle deformation
Change in tablet microstructure due to rearrangement of solid material in the amorphous state
Change in tablet microstructure due to polymorphic transformations

mechanisms responsible for post-compaction changes in tablet strength, studies on simple, one-component powders have been conducted (Table 27.5). The suggested mechanisms are probably active also in tablets formed from complete formulations of several components.

During storage at a fairly high relative humidity, tablets can be softer and their tensile strength reduced. With increased relative humidity, the state of water adsorbed at the solid surface can change from an adsorbed gas to a liquid, i.e. water condenses in the tablet pores. Furthermore, if the solid material is freely soluble in water, it can dissolve. Both the presence of condensed water in the pores and the dissolution of a substance in the condensed water can drastically decrease tablet strength and eventually lead to the collapse of the whole tablet. However, the dissolution of a freely soluble substance in condensed pore water can also give an increase in tablet strength if the water is allowed to evaporate owing to a change in temperature or relative humidity. The result of this evaporation can be a crystallization of solid material, with the subsequent formation of solid bridges between particles in the tablet and increased tablet strength.

In addition to the mechanisms involving the presence of condensed pore water, several other mechanisms causing an increase in tablet strength during storage at a relative humidity at which condensation of water is unlikely to occur, have been proposed. One such mechanism is a continuing viscous deformation of particles after the compaction process is completed. This phenomenon is referred to as ***stress relaxation*** of tablets. The increase in tablet strength can be significant with no or minor detectable changes in its physical structure. However, viscous deformation of small parts of particles might change the microstructure of the tablet in terms of the relative orientation of particle surfaces, and the geometry of the interparticle voids, and thus affect the resistance to fracturing of a tablet. A characteristic feature for stress relaxation changes is that the tablet strength changes occur for a limited time in connection with the compaction phase.

Explanations for observed changes in tablet strength due to the presence of amorphous material in tablets have been presented. If the amorphous substance absorbs or desorbs moisture, the mechanical strength of the tablet can change. This is probably related to an effect of absorbed moisture on the mechanical properties of the amorphous material. If the uptake of moisture allows the amorphous material to change from a glassy to a rubbery state, the amorphous phase may crystallize. Such crystallization can subsequently affect – normally increase – tablet strength.

Another mechanistic explanation for a storage-related increase in tablet strength which involves amorphous material is a process described as a restructuring of parts of the pore system due to a rearrangement of solid material. It has been reported, for example, that a marked increase in tablet strength during storage occurred in parallel with a change to a more open pore structure. Moreover, these changes do not necessarily occur immediately after compaction, but can be initiated by exposure of the tablet to humid air for a certain period after compaction (Fig. 27.35). The restructuring of the pore system might be due to a diffusion-like transport of molecules or ions at the surface of particles, followed by a localization of material in zones where particle surfaces are close to each other. Such a mobility of molecules in the solid state has been shown by an amorphous phase.

Finally, an increase in tablet strength during storage has also been explained by a change in the crystal structure of particles, from a less to a more stable crystal form, i.e. a polymorphic transformation.

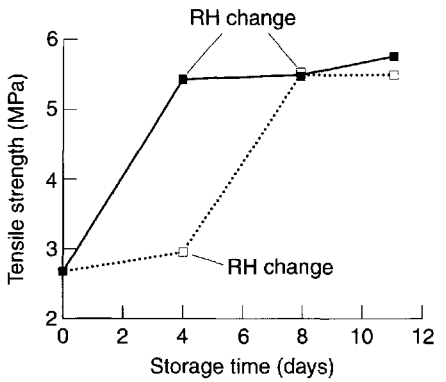


Fig. 27.35 The tensile strength of sodium chloride tablets as a function of storage time for tablets stored at different relative humidities. Open symbols: tablets stored 4 days at low, 4 days at high and 3 days at low relative humidity. Closed symbols: tablets stored 4 days at high, 4 days at low and 3 days at high relative humidity. (From Eriksson, M. and Alderborn, G. Int. J. Pharm. 109, 59, 1994.)

RELATIONSHIPS BETWEEN MATERIAL PROPERTIES AND TABLET STRENGTH

Factors of importance for powder compactability

A number of empirical studies exists in the pharmaceutical literature with the aim of mapping factors that affect the structure of a tablet and its mechanical strength, i.e. tensile strength, resistance towards attrition and capping tendencies. These factors can be classified into three group: material and formulation factors; processing factors (choice of tablet machine and operation conditions); and environmental factors (relative humidity etc.).

Of special importance from a formulation perspective are the physical and mechanical properties of the particles used in the formulation and how these particles are combined in granulation and mixing steps. Relationships for powders consisting both of one component and of two components, such as a filler and a lubricant or a dry binder, have been discussed in this context.

The compaction of solid particles

It is often assumed that the evolution of the interparticulate structure of a tablet, in terms of bonds between particles and the pores between the particles, will be significant for the mechanical strength of the tablet. Thus, the material-related factors that control the evolution of the microstructure of the tablet have been discussed as important factors for the compactability of a powder. In this context, the compression behaviour and the original dimensions of the particles have received special interest.

As discussed above, the degree of fragmentation and permanent deformation that particles undergo during compression are significant for tablet structure and strength. It has been suggested that both fragmentation and deformation are strength-producing compression mechanisms. The significance of particle fragmentation has been considered to be related to the formation of small particles which constitute the tablet, with the consequence that a large number of contact sites between particles at which bonds can be formed will be developed. The significance of permanent deformation has been explained in terms of an effect on the area of contact of the interparticulate contact sites, with a subse-

quent increased bonding force. The relative importance of these mechanisms for the bonding between particles in a tablet and the resistance of a tablet towards fracturing has, not however, been fully clarified. Concerning elastic deformation, which is recoverable, this is considered as a disruptive rather than a bond-forming mechanism. Poor compactability, in terms of low tablet strength and capping/lamination, has been attributed to elastic properties of the solid. A summary of proposed advantages and disadvantages of the different particle compression mechanisms for the ability of the particles to form tablets is given in Table 27.6.

It is sometimes considered that one of the most important properties of particles for the mechanical strength of a tablet is their size before compaction. A number of empirical relationships between particle dimensions before compaction and the mechanical strength of the resultant tablet can thus be found in the literature. As a rule, it is normally assumed that a smaller original particle size increases tablet strength. However, it is also suggested that the effect of original particle size is in relative terms limited for powder compactability, with the possible exception of very small (i.e. micronized) particles. Reported data show, however, that different and sometimes complex relationships between particle size and tablet strength can be obtained, with maximum or minimum tablet strength values. Complex relationships might be associated with a change in the shape, structure (such as the formation of aggregates) or degree of disorder of the particles with particle size. It seems also that increased compaction pressure stresses the relationship between original particle size and tablet strength in absolute terms.

Expressions quantifying the relationship between tablet strength and original particle size have been presented in the literature, such as the following:

$$F = K d^{-a}$$

where F is the force (N) needed to break the compact, d is the diameter (m) of the particle, and, K and a are constants. The expression thus describes the general assumption that tablet strength increases with a reduced original particle size. The compactability of sodium chloride and hexamine is described by this equation (Fig. 27.36).

Some studies have specifically reported on the effect of original particle shape on tablet strength. The results indicate that, for particles that fragment to a limited degree during compression, an increased particle irregularity improved their compactability. However, for particles that fragmented markedly during compression the original shape of the particles did not affect tablet strength. Moreover, an increased compaction pressure increased the absolute difference in strength of compacts of different original particle shape. Thus, the shape characteristics of particles that fragment markedly during compression seem not to affect the microstructure and the tensile strength of tablets, but the converse applies for particles that showed limited fragmentation.

Table 27.6 Proposed advantages and disadvantages of the different compression mechanisms in relation to the tablet-forming ability of the powder

Compression mechanism	Advantages	Disadvantages	Others
Fragmentation	No effect of particle shape Low sensitivity to additives Strain-rate insensitive	May cause fracturing of tablets (capping etc.)	Bond-forming ability (and tablet strength) dependent on degree of particle fragmentation
Plastic deformation	Resistant towards fracturing of tablets (capping etc.) Strain-rate insensitive	Sensitive to additives and variations in original particle shape	Bond-forming ability (and tablet strength) dependent on degree of particle deformation
Elastic deformation	–	May cause fracturing of tablets (capping etc.)	–
Time-dependent deformation	–	Strain rate sensitive Prone to change tablet strength after compaction due to stress relaxation	–

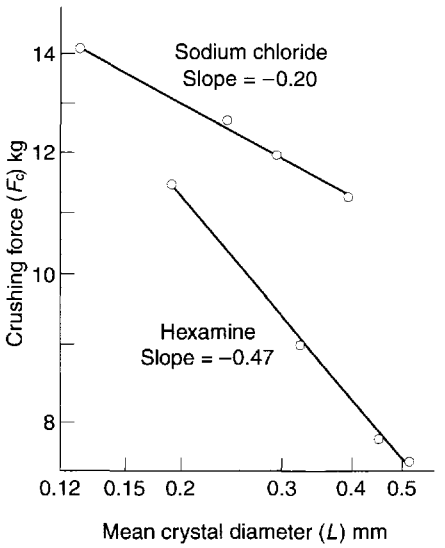


Fig. 27.36 The relationship (log-log scales) between the force needed to break the tablet and the original diameter of the particles of sodium chloride and hexamine. (From Shotton, E. and Ganderton, D. J. Pharm. Pharmacol. Suppl. 13, 144T, 1961.

The compaction of granules

The rationale for granulating a powder mixture before tabletting has been discussed above, one reason being to ensure good compactability. When granules are compacted, the mechanical characteristics of the primary particles will probably affect the compactability of the mass. For example, it is a common experience that capping-prone material will show capping tendencies in the granulated form of the substance also. However, the design of the granulation process, such as the method of granulation, will also affect the compactability of the granules. Such process conditions will control the physical properties of the aggregates formed, e.g. in terms of intragranular binder distribution and granule porosity.

Tablets formed from granules can in physical terms be described as granules bonded together by intergranular bonds. When subjected to a load, tablets formed from granules often fail because of breakage of these bonds. Hence, the bonding force of the intergranular bonds and the structure of the intergranular pores will be significant for the tensile strength of the tablets. The evolution of the intergranular tablet microstructure during compression is affected by the physical properties of the granules before compression, as well as their composition.

Thus, in order to engineer granules in terms of their compactability, two main factors can be controlled:

- The composition of the granules (e.g. choice of filler and binder);
- The physical properties of the granules (e.g. porosity and mechanical strength).

In terms of the physical properties of granules, porosity, compression shear strength and shape are significant properties that influence compactability. In general terms, increased porosity, decreased compression shear strength and increased irregularity (Fig. 27.37) will increase the compactability of the granules. As discussed above, pharmaceutical granules seem to fragment to only a limited degree during compression. The importance of these granule properties for compactability has thus been discussed in terms of a sequential relationship between the original physical character of the granules, the degree of deformation they undergo during compression, and the area of contact and the geometry of the intergranular pores of the formed tablet. The formation of large intergranular areas of contact and a closed pore system promotes a high tablet strength.

Traditionally, the most important means to control the compactability of granules has been to add a binder to the powder to be granulated. This is normally done by adding the binder in a dissolved form, thereby creating binder–substrate granules. An increased amount of binder can correspond to an increased compactability, but this is not a general rule. The importance of the presence of a binder for the compactability of such granules can be explained in two ways. First, it has been suggested that intergranular bonds that involve binder-coated granule surfaces can be described as comparatively strong,

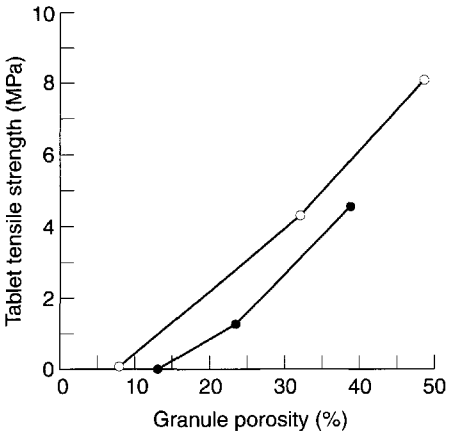


Fig. 27.37 The tensile strength of tablets formed from granules of a series of porosities and of two different shapes. Open symbol: irregular granules. Closed symbol: nearly spherical granules. (From Johansson, B. and Alderborn, G. Unpublished data.)

i.e. difficult to break. Secondly, binders are often comparatively deformable substances, which can reduce the compression shear strength of the whole granule and thus facilitate the deformation of the granules during compression. An increased degree of granule deformation is sometimes proposed to increase the compactability of the granules. Thus, the binder might have a double role in the compactability of granules, i.e. increase granule deformation and increase bond strength. Except for the presence of a binder in the granules, the combination of fillers in terms of the hardness and dimensions of the particles can affect the compression shear strength and hence the deformation properties of the granules during compression.

In the preparation of binder–substrate granules the intention is normally to spread out the binder homogenously within the granules, i.e. all substrate particles are more or less covered with a layer of binder. However, it is possible that the binder will be concentrated at different regions within the granules, e.g. due to solute migration during drying. The question of the importance of a relatively homogenous distribution versus a peripheral localization of the binder, i.e. concentration at the granule surface, has been addressed in the literature. It has been argued that a peripheral localization of the binder in the granules before compression should be advantageous, as the binder can thereby be used most effectively for the formation of intergranular bonds. However, the opposite has also been suggested, i.e. a homogenous binder distribution is advantageous for the compactability of granules. This observation was explained by assuming that, owing to extensive deformation and some attrition of granules during compression, new extragranular surfaces will be formed originating from the interior of the granule. When the binder is distributed homogenously, such compression-formed surfaces will show a high capacity for bonding.

Figure 27.38 gives an overview of the physical granule properties that affect the compactability of granules.

The compaction of binary mixtures

Most of the fundamental work on powder compaction has been carried out on one-component powders. It is, however, of obvious interest to derive knowledge that enables the prediction of the compactability of mixtures of powders from the information from the compaction behaviour of the individual components. In this context, powder mixtures of two components, i.e. binary mixtures, have been the system of choice in pharmaceutical studies. Binary mixtures can be of two types: simple physical mixtures, i.e. nearly randomized mixtures of particles, and interactive (ordered) mixtures. Most of the studies in this context are empirical, although models for the compaction of binary powder mixtures have been derived.

Concerning simple binary mixtures, the importance of the relative proportions of the ingredients has been studied in relation to the compactability of

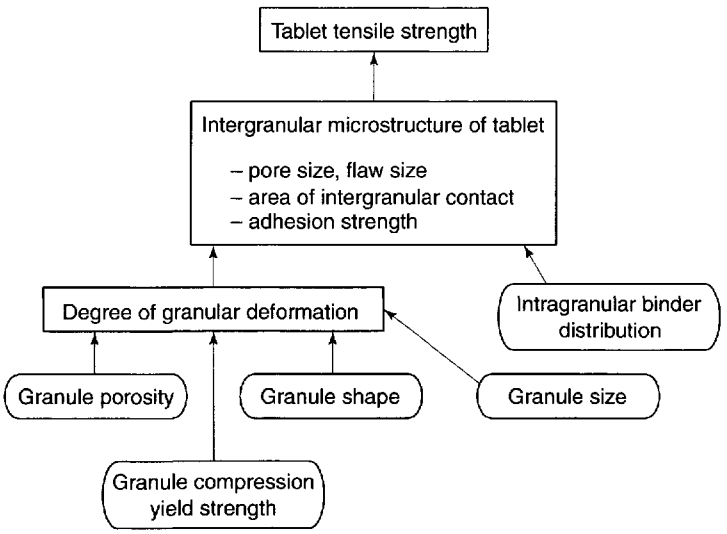


Fig. 27.38 Overview of proposed physical granule properties of importance for the compactability of granules.

the respective single component. The mixture can show a change in compactability, as can be predicted proportionally from the compactability of the single powders, but deviations from such a simple linear relationship, both positively and negatively, have also been reported. Such non-linear behaviour has been explained in terms of differences between the components in their mechanical and adhesive properties.

Interactive mixtures, especially their compactability after the admixture of lubricants and dry binders, have been the subject of study. Concerning the tablet strength-reducing effect of a lubricant mixed with solid particles, it depends on the surface coverage of the lubricant film obtained during mixing, on the compaction properties of the lubricant per se, and on the compression behaviour of the substrate particles. Lubricant sensitivity, also referred to as dilution capacity, seems to be strongly related to the fragmentation propensity of the substrate particles, as discussed earlier.

Concerning the tablet-strength increasing effect of a dry binder mixed with solid particles, similar factors seem to control the compactability of the dry binder mixture as for the lubricant mixture, i.e. the degree of surface coverage of the substrate particle, the binding capacity and deformability of the dry binder, and the fragmentation propensity of the substrate particles (Fig. 27.39).

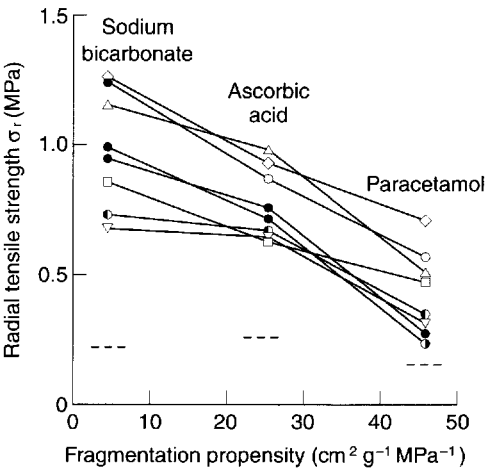


Fig. 27.39 The tensile strength of tablets formed from three substrate substances of different fragmentation propensities in binary mixtures with some fine particulate dry binders i.e. microcrystalline cellulose, methylcellulose and polyvinylpyrrolidone of different particle size. The dotted lines represent the tensile strength of tablets formed from the single substrate substances. (From Nyström, C. and Glazer, M. Int. J. Pharm. 23, 255, 1985.)

The dilution capacity of interactive mixtures between granules and lubricants or dry binders seems to be related to the degree of deformation the granules undergo during compression, i.e. a high degree of deformation will give a lower sensitivity to a lubricant but also a less positive effect of a dry binder.

REFERENCES

Adams, M.J., Mullier, M.A., Seville, J.P.K. (1994) Agglomerate strength measurement using a uniaxial confined compression test. *Powder Technol.*, **78**, 5.

Eriksson, M. and Alderborn, G. (1995) The effect of particle fragmentation and deformation on the interparticulate bond formation process during powder compaction. *Pharm. Res.* **12**, 1031.

Hiestand, E.N. (1996) Rationale for and the measurement of tableting indices. In: Pharmaceutical Powder Compaction Technology (eds. Alderborn, G. and Nyström, C.), p. 219, Marcel Dekker Inc., New York.

Kendall, K. (1988) Agglomerate strength. *Powder Metall.*, **31**, 28.

Leuenberger, H. (1982) The compressibility and compactability of powder systems. *Int. J. Pharm.* **12**, 41.

Roberts, R. and Rowe, R. (1985) The effect of punch velocity of the compaction of a variety of materials. *J. Pharm. Pharmacol.*, **37**, 377.

Sebhatu, T. and Alderborn, G. (1999) Relationships between the effective interparticulate contact area and the tensile strength of tablets of amorphous and crystalline lactose of varying particle size. *Eur. J. Pharm. Sci.* **8**, 235.

BIBLIOGRAPHY

Alderborn, G. and Nyström, C. (1996) *Pharmaceutical Powder Compaction* Technology. Marcel Dekker Inc., New York.

Alderborn, G., Pasanen, K., and Nyström, C. (1985) Studies on direct compression of tablets. XI. Characterization of particle fragmentation during compaction by permeametry measurements of tablets, *Int. J. Pharm.*, **23**, 79.

Al-Nasassrah, M.A., Podczeck, F. and Newton, J.M. (1998) The effect of an increase in chain length on the mechanical properties of polyethylene glycols. *Eur. J. Pharm. Biopharm.* **46**, 31.

Armstrong, N.A. (1982) Causes of tablet compression problems. *Mfg Chem.* October 62.

Banakar, U.V. (1992) *Pharmaceutical Dissolution Testing.* Marcel Dekker, Inc., New York.

Chien, Y.W. (1992) *Novel Drug Delivery Systems*, 2nd edition. Marcel Dekker, Inc., New York.

Chulia, D., Deleuil, M. and Pourcelot, Y. (1991) *Powder Technology and Pharmaceutical Processes* (Handbook of Powder Technology, vol. 9). Elsevier, Amsterdam.

Duberg,. M. and Nyström, C. (1986) Studies on direct compression of tablets. XVII. Porosity–pressure curves for the characterization of volume reduction mechanisms in powder compression. *Powder Technol.* **46**, 67.

Eriksson, M. and Alderborn, G. (1994) Mechanisms of post-compaction changes in tensile strength of sodium chloride compacts prepared from particles of different dimensions. *Int. J. Pharm.* **109**, 59.

European Pharmacopoeia 3rd edition 1997. Council of Europe, Strasbourg.

Hölzer, A.H. and Sjögren, J. (1981) Evaluation of some lubricants by the comparison of friction coefficients and tablet properties. *Acta Pharm. Suec.* **18**, 139.

Lieberman, H.A., Lachman, L. and Schwartz, J.B. (1989) *Pharmaceutical Dosage Forms: Tablets* (2nd edition, volumes 1–3). Marcel Dekker, Inc, New York.

Nyström, C. and Glazer, M. (1985) Studies on direct compression of tablets. XIII. The effect of some dry binders on the tablet strength of compounds with different fragmentation propensity. *Int. J. Pharm.* **23**, 255.

Podczeck, F. (1998) *Particle–particle Adhesion in Pharmaceutical Powder Handling*. Imperial College Press, London.

Sandell, E. (1993) *Industrial Aspects of Pharmaceutics*. Swedish Pharmaceutical Press, Stockholm.

Stanley-Wood, N.G. (1983) *Enlargement and Compaction of Particulate Solids*. Butterworths, London.

28
Coating of tablets and multiparticulates

John Hogan

CHAPTER CONTENTS

DEFINITION

Tablet coating is the application of a coating material to the exterior of a tablet with the intention of conferring benefits and properties to the dosage form over the uncoated variety.

In its widest sense the technology is also applicable to multiparticulate systems intended for modified-release applications. To a much lesser extent coatings may also be applied to hard-shell and soft elastic capsules.

Types of tablet coating

Three main types are in use:

- Film coating
- Sugar coating
- Press coating.

Of these, film coating is the major technique: virtually all new coated products introduced on to the market are film coated. Sugar coating is the more traditional technology and has seen no real developments in recent years. As a proportion of the total output of coated tablets on a global basis, though, it is still of some economic importance.

Reasons for coating tablets

The reasons why tablets are coated are varied. The major ones can be summarized as follows:

1. Ingredients may need protection from the environment, particularly light and moisture.
2. Many drugs have a bitter or otherwise unpleasant taste: coating is an efficient way to mask such tastes. Tablets that are coated are also somewhat easier to swallow than uncoated tablets.

3. Coloured coatings also mask any batch differences in the appearance of raw materials and hence allay patient concern over tablets of differing appearance.

Factors 2 and 3 aid patient compliance with dosage schemes.

4. Coatings may be optimized with respect to colouration and gloss to aid in their sales appeal or to reinforce a marketing brand identification.
5. Coloured coatings aid in the rapid identification of product by the manufacturer, the dispensing pharmacist and the patient.
6. Coating tablets facilitates their handling on high-speed automatic filling and packaging equipment. Very often coating confers an added mechanical strength to the tablet core. Cross-contamination is also reduced in the manufacturing plant, as 'dusting' from tablets is eliminated by coating.
7. Functional film coatings are used to impart enteric or controlled-release properties to the coated tablet or, more usually, to coated multiparticulates (see later).

FILM COATING

This is the more modern and generally used technology in tablet coating. Nearly all newly launched coated products are film coated rather than sugar coated, for the reasons given in Table 28.1.

Process description

Film coating involves the deposition, usually by a spray method, of a thin film of polymer surrounding the tablet core. It is possible to use conventional panning equipment, but more usually specialized equipment is employed to take advantage of the fast coating times and high degree of automation possible.

The coating liquid (solution or suspension) contains a polymer in a suitable liquid medium together with other ingredients such as pigments and plasticizers. This solution is sprayed on to a rotating, mixed tablet bed or fluid bed. The drying conditions permit the removal of the solvent so as to leave a thin deposition of coating material around each tablet core.

Coating suspension formulation

Typically this comprises:

- Polymer
- Plasticizer
- Colourants
- Solvent.

Ideal characteristics of a film coating polymer

Solubility

For conventional film coating the polymer should have good solubility in aqueous fluids to facilitate the dissolution of the active ingredient from the finished dosage form. However, where a modified-release

Table 28.1 Major differences between sugar and film coating

Features	Sugar coating	Film coating
Tablets		
Appearance	Rounded with high degree of polish	Retains contour of original core. Usually not as shiny as sugar coat types
Weight increase due to coating materials	30–50%	2–3%
Logo or 'break' lines	Not possible	Possible
Other solid dosage forms	Coating possible but little industrial importance	Coating of multiparticulates very important in modified release forms
Process		
Stages	Multistage process	Usually single stage
Typical batch coating time	Eight hours, but easily longer	1.5–2 hours
Functional coatings	Not usually possible apart from enteric coating	Easily adaptable for controlled release

action is required then a polymer system of low water solubility or permeability will be chosen.

Viscosity

In general, polymers should have a low viscosity for a given concentration. This will permit the easy, trouble-free spraying of their solutions in industrial film coating equipment.

Permeability

Film coating can be used to optimize the shelf-life of a tablet preparation, as some polymers are efficient barriers against the permeability of water vapour or other atmospheric gases. These properties vary widely between the individual polymers.

Mechanical properties

The particular polymer chosen for a film coat formulation must be one with adequate strength to withstand the impact and abrasion encountered in normal handling. Insufficient coating strength will be demonstrated by the development of cracks and other imperfections in the coating.

It should be mentioned that the polymer chosen must also comply with the relevant regulatory and pharmacopoeial requirements current in the intended marketing area.

Types of polymer available

Cellulose derivatives

Most are substituted ethers of cellulose.

Hydroxypropyl methylcellulose is the most widely used of the cellulosic polymers (Fig. 28.1). It is soluble in aqueous media and forms films which are mechanically tough and relatively easy to apply. The resultant film can be clear or coloured with permitted pigments. The polymer is the subject of monographs in the major international pharmacopoeia.

Other cellulosic derivatives used in film coating include methylcellulose and hydroxypropyl cellulose.

Methacrylate amino ester copolymers

Basically these polymers are insoluble in water below pH 4, but in neutral or alkaline media the films achieve solubility by swelling and increased permeability. For simple formulations the disintegration of the coating can be optimized by the incorporation of water-soluble materials and also by starches. Chemically an example is the polymer poly(butylmethacrylate) (2-dimethylaminoethyl) methacrylate methylmethacrylate.

For coatings designed to confer a modified-release aspect to the final dosage form, more water-insoluble polymers are used. These include ethylcellulose and the ammonio methacrylate copolymers. Yet another group of polymers is designed to provide an enteric protection to the dosage form. This effect is achieved by a pH selectivity of the polymer where it is insoluble at the low pH environment of the stomach yet becomes soluble as the higher pH of the duodenum and the distal portion of the gastrointestinal system is reached.

Aqueous polymer dispersions

Industrially, specialized dispersions of water-insoluble polymers such as ethylcellulose and ammonio methacrylate copolymers for use in aqueous media are frequently encountered in the coating of beads and granules for use in modified-release preparations (Zhang et al 1989). The advantage of these materials is that they permit the aqueous processing of otherwise water-insoluble polymers, with the consequent benefits of this method of processing (see 'Solvents', below).

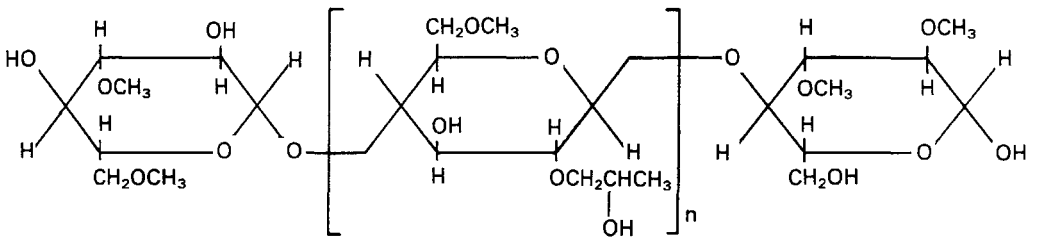


Fig. 28.1 Hydroxypropyl methylcellulose.

Plasticizers

Plasticizers are generally added to film coating formulations to modify the physical properties of the polymer to make it more usable. One important property is their ability to decrease film brittleness. It is generally accepted that the mechanism by which polymers exert their action is for them to interpose themselves on a molecular scale between the polymer strands. In doing so they permit these strands to move more freely and allow the polymer to act in a more pliable fashion.

Examples of plasticizers are:

- polyols, such as polyethylene glycol 400
- organic esters, such as diethyl phthalate
- oils/glycerides, such as fractionated coconut oil.

In general, only water-miscible plasticizers can be used for aqueous-based spray systems.

Colourants

Any permitted colourants in a film coat formula are invariably water-insoluble colours (pigments). Pigments have certain advantages over water-soluble colours: they tend to be more chemically stable towards light, provide better opacity and covering power, and optimize the impermeability of a given film to water vapour.

Examples of colourants are:

- iron oxide pigments
- titanium dioxide
- aluminium Lakes.

Solvents

After the early development of film coating in the 1950s the polymers used were invariably dissolved in an organic solvent. Modern techniques now rely on water as a solvent because of the significant drawbacks that readily became apparent with the use of organic solvents. The disadvantages of organic solvents for the process can be listed below (see also Hogan 1982).

1. *Environmental*: the venting of untreated organic solvent vapour into the atmosphere is ecologically unacceptable, and efficient solvent vapour removal from gaseous effluent is expensive.
2. *Safety*: organic solvents provide explosion, fire and toxic hazards to plant operators.
3. *Financial*: the use of organic solvents necessitates the building of flame- and explosion-proof facilities. Ingredient cost is also comparatively high, and the associated costs of storage and quality control must also be taken into consideration.
4. *Solvent residues*: for a given process the amount of residual organic solvent in the film must be investigated. With increasing regulatory pressure this will become an area for additional control in the future.

Process details

The vast majority of film-coated tablets are produced by a process which involves the atomization (spraying) of the coating solution or suspension on to a bed of tablets.

Some examples of equipment suitable for film coating include:

- Accela Cota – Manesty Machines, Liverpool, UK
- Hi-Coater – Freund Company, Japan
- Driacoater – Driam Metallprodukt GmbH, Germany
- HTF/150 – GS, Italy
- IDA – Dumoulin, France.

Examples of units that function on a fluidized-bed principle include:

- Aeromatic-Fielder, Switzerland and UK
- Glatt AG, Switzerland and Germany

Figure 28.2 shows one of the most widely used pieces of equipment for film coating, the Accela Cota.

Basic process requirements for film coating

These fundamental requirements are more or less independent of the actual type of equipment being used and include:

1. adequate means of atomizing the spray liquid for application to the tablet cores;
2. adequate mixing and agitation of the tablet bed. Spray coating relies upon each core passing through the area of spraying. This is distinct from sugar coating, where each application is spread from tablet to tablet prior to drying;
3. sufficient heat input in the form of drying air to provide the latent heat of evaporation of the solvent. This is particularly important with aqueous-based spraying;
4. good exhaust facilities to remove dust- and solvent-laden air.

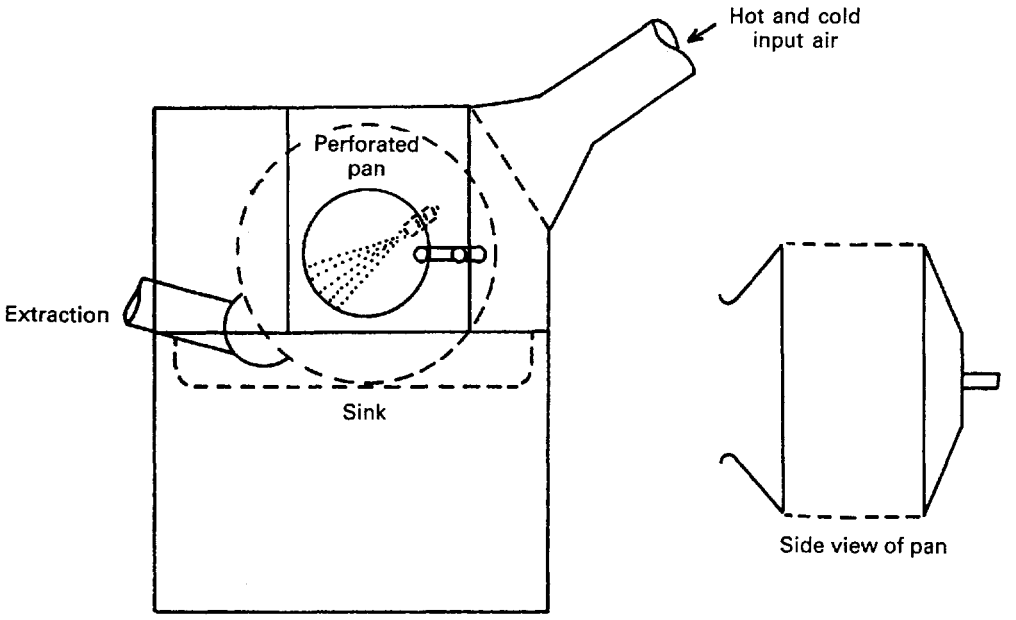


Fig. 28.2 Accela Cota.

Ideal characteristics of film-coated tablets

Film-coated tablets should display an even coverage of film and colour. There should be no abrasion of tablet edges or crowns. Logos and break lines should be distinct and not filled in. The tablet must also be compliant with finished product specifications and any relevant compendial requirements.

Coating faults

These arise from two distinct causes:

1. *Processing*: for example, inadequate drying conditions will permit coating previously deposited on the tablet surface to stick against neighbouring tablets. When parted, this will reveal the original core surface underneath.
2. *Formulation faults*: film cracking or 'bridging' of break lines are examples of this type. After taking due account of the mechanical properties of the film, reformulation will almost certainly be successful in overcoming the problem (Rowe 1981).

SUGAR COATING

Sugar coating may be considered the traditional method of coating tablets. It involves the successive application of sucrose-based solutions to tablet cores in suitable coating equipment. Conventional panning equipment with manual application of syrup has been extensively used, although more specialized equipment and automated methods are now making an impact on the process. A comparison between sugar coating and film coating has been given in Table 28.1.

Stages involved in the production of sugar-coated tablets

Sugar coating is a multistage process and can be divided into the following steps:

1. Sealing of the tablet cores
2. Subcoating
3. Smoothing
4. Colouring
5. Polishing
6. Printing.

Initially the tablet cores to be sugar coated are sealed against the entry of water by the application of a water-impermeable polymer. Shellac has traditionally been used for this purpose and is indeed still used a great deal today, although more reliable materials, such as cellulose acetate phthalate and polyvinyl acetate phthalate, also find favour.

To attain the typically rounded profile of a sugar-coated tablet the sealed tablet core must be built up

to gain the desired profile. This process of subcoating is usually performed by adding bulking agents such as calcium carbonate or talc to the applied sucrose solutions. A gum such as acacia is also added to the applied suspension.

After the correct profile has been obtained the subcoated tablets will almost certainly have a rough surface, which will have to be made smooth before the next stage can be commenced. This is accomplished by the application of a few coats of sucrose syrup.

Nearly all sugar-coated tablets are coloured, as aesthetic appearance is usually considered to be of great importance with this dosage form. The pigments used are those permitted by the national legislation of the country where the products are to be marketed.

After the colour-coating stage the tablets will require a separate polishing stage for them to acquire an acceptable appearance. Several methods can be used, but commonly beeswax and carnauba wax are used in the process.

To facilitate identification sugar-coated tablets are usually printed with a manufacturer's logo or code. The use of indented monograms for this purpose, as for film-coated tablets, would not be feasible as the considerable thickness of sugar coating would obliterate any core markings. The printing process used is an offset gravure in conjunction with special edible inks, although the inkjet process is starting to make an impact.

Process details

Typically tablets are sugar coated by a panning technique. The simplest form would be a traditional sugar-coating pan with a supply of drying air (preferably of variable temperature and thermostatically controlled) and a fan-assisted extract to remove dust- and moisture-laden air.

Methods of applying the coating syrup include manually using a ladle, and, automatic control. In modern equipment some form of automatic control is available for the application of coating syrups.

In general, the equipment listed under film coating can, with suitable modification, be used for sugar-coating techniques.

Ideal characteristics of sugar-coated tablets

First the tablets must comply with finished product specifications and any appropriate compendial requirements. Sugar-coated tablets should ideally be of a perfectly smooth rounded contour with even colour coverage. Most manufacturers take advantage of the aesthetic appeal of a sugar-coated tablet and polish to a high gloss. Any printing should be distinct, with no smudging or broken print.

Coating faults

These are usually associated with process defects, such as splitting of the coat on storage, caused by inadequate drying during the coating application.

PRESS COATING

The technology of press coating differs radically from the previously described film- and sugar-coating techniques. Press coating involves the compaction of granular material around an already preformed core (Fig. 28.3) using compressing equipment similar to that used for the core itself, e.g. Manesty Drycota.

Today press coating is used mainly to separate chemically incompatible materials, one or more being placed in the core and the other(s) in the coating layer. However, there is still an interface of contact left between the two layers. In cases where even this is important then the process of press coating can be taken one stage further. It is possible to apply two press coatings to a tablet core using suitable equipment, e.g. Manesty Bicota. This equipment produces press-coated tablets with perfect separation between active core and coating, as the two can be separated by an inert middle layer.

The formulation and processing of the coating layer requires some care. Large or irregularly sized agglomerates of granules will cause the core to tilt in the second die used for compression of the coating. Thus there is the possibility of an incomplete coating, with the core being visible at the tablet surface.

The disadvantages of the process arise from the relative complexities of the mechanism used in the compressing equipment.

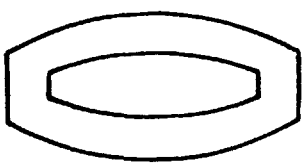


Fig. 28.3 Compaction of granular material around an already preformed core.

FUNCTIONAL COATINGS

All the coatings described above have been designed as a taste mask, as an identification aid, or indeed for many of the reasons previously discussed for coating tablets. There are, however, tablet coatings that perform a pharmaceutical function, such as conferring controlled or enteric release on the dosage form.

Controlled-release coatings

Film coating provides an extremely effective way of conferring a controlled-release aspect to a tablet or, more usually, a multiparticulate system. After coating these particles are filled into hard gelatin shells, or occasionally compressed directly into tablets by a process which permits minimal rupture of the applied film. The coatings involved use polymers with restricted water solubility or permeability, and include ethylcellulose and modified acrylate derivatives.

Multiparticulates, commonly referred to as 'pellets' or 'beads', find favour over conventional non-disintegrating tablets for controlled release use, owing to a number of factors:

1. Their small size (typically 0.7–2.00 mm) allows them to pass through the constricted pyloric sphincter and distribute themselves along the gastrointestinal tract. This tends to overcome the disadvantage that whole tablets have of a rather irregular passage through the gastrointestinal tract and consequent irregular absorption
2. Whole, non-disintegrating tablets can be liable to lodge in restrictions within the gastrointestinal tract, and this can lead to ulcerative damage to the gastric mucosa as the drug solution is leached out from the tablet. Because of their small size, this is not a problem with multiparticulates.
3. Should an individual bead or pellet fail and release all of its contents at once the patient would not be exposed to any undue risk. This is certainly not the case if a non-disintegrating tablet failed, when the consequences would potentially be serious.

Types of multiparticulate

Extruded/spheronized granulates

These are produced in modified granulating equipment, with the drug granulation extruded through a mesh or other device under pressure to form small granulates which are subsequently spheronized.

Non-pareils

These are sucrose spheres which are coated with the drug plus an adhesive yet water-soluble polymer (Fig 28.4). After their formation and any necessary intermediate steps such as drying, they may be coated with the controlled-release coating.

Mechanisms of drug release from multiparticulates

Subsequent to the release of the coated pellets from the hard-shell capsule or tablet, the drug is released in a predetermined fashion with respect to time. The mechanisms described below postulate how this may be achieved (for further detail, see Chapter 20).

Diffusion

On contact with the aqueous fluids of the gastrointestinal tract, water will enter the interior of the particle by diffusion. Dissolution of the drug will occur and drug solution will diffuse across the controlled-release coat to the exterior. The kinetics of the process will depend upon which is the rate-controlling step, the dissolution of the drug or the diffusion of the drug solution through the coating.

Erosion

Some coatings can be designed to erode gradually with time, thereby releasing the drug contained within the pellet.

Osmosis

In allowing water to enter under the right circumstances, an osmotic pressure can be built up within

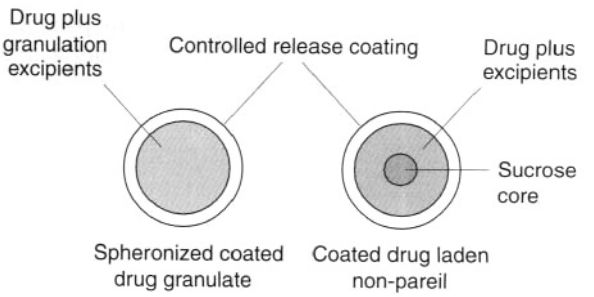


Fig. 28.4 Differences between multiparticulate types.

the interior of the pellet. Should the coating contain micro-imperfections and cracks the drug solution will be forced out of the pellet to the exterior.

Enteric coating

This technique is used to protect the tablet core from disintegration in the acid environment of the stomach for one or more of the following reasons:

1. Prevention of acid attack on active constituents unstable at low pH;
2. To protect the stomach from the irritant effect of certain drugs;
3. To facilitate absorption of a drug that is preferentially absorbed distal to the stomach.

The following polymers are among those commonly used for the purposes of enteric coating:

- Cellulose acetate phthalate
- Polyvinyl acetate phthalate
- Suitable acrylic derivatives.

Because they possess free carboxylic acid groups on the polymer backbone, they exhibit a differential pH solubility profile. They are almost insoluble in aqueous media at low pH, but as the pH rises they experience a sharp, well defined increase in solubility at a specific pH, e.g. pH 5.2 for cellulose acetate phthalate.

Enteric coating is possible using both sugar- and film-coating techniques.

Enteric film coating

The enteric polymers listed are capable of forming a direct film in a film-coating process. Sufficient weight of enteric polymer must be used to ensure an efficient enteric effect. This is normally two or three times that required for a simple film coating.

Enteric sugar coating

The sealing coat is modified to comprise one of the enteric polymers in sufficient quantity to pass the enteric test for disintegration. The subcoating and subsequent coating steps are then as for conventional sugar coating.

STANDARDS FOR COATED TABLETS

The European Pharmacopoeia has similar requirements for coated and uncoated tablets, the differences being:

1. Film-coated tablets must comply with the uniformity of mass test ***unless otherwise justified and authorized.***
2. Film-coated tablets comply with the disintegration test for uncoated tablets except that the apparatus is operated for 30 minutes. The requirement for coated tablets other than film coated is modified to include a 60-minute operating time. Furthermore, the test may be repeated using 0.1 N HCl in the event that any tablets fail to disintegrate in the presence of water.

REFERENCES

Hogan, J.E. (1982) Solvent versus aqueous film coating. *Int. J. Pharm. Tech. Prod. Mfr*, **3**, 7.

Rowe, R.C. (1981) The cracking of film coatings on film coated tablets – a theoretical approach with practical implications. *J. Pharm. Pharmacol.*, **33**, 423.

Zhang, G., Schwartz, J.B., Schnaare, R.L. (1989) 6th Int. Symp. Controlled Release of Bioactive Materials, Chicago.

BIBLIOGRAPHY

Aulton, M.E., Abdul-Razzak, M.H., Hogan. J.E. (1981) The mechanical properties of hydroxypropyl methycellulose films derived from aqueous systems. Part 1: The influence of plasticisers. *Drug Dev. Ind. Pharm.*, **7**, 649.

Cole, G., Hogan, J.E., Aulton, M., eds (1995) *Pharmaceutical Coating Technology*. Taylor & Francis.

Hogan, J.E. (1983) Additive effects on aqueous film coatings. *Mfg. Chem.*, **54**, 43.

Porter, S.C. (1981) Tablet coating. *Drug Cosmetic Ind.*, May 46, June 44, Aug. 40, Sept. 50.

Rowe, R.C. (1983) Coating defects – causes and cures. *Mfg Chem.*, **54**, 49.

Melia, C.D., Washington, N., Wilson, C.C. (1994). *Multiparticulate Controlled Release Oral Dosage Forms*. Scottish Academic Press.

European Pharmacopoeia, 3rd edn. Suppl. 2000. Council of Europe, Strasbourg.

29
Hard gelatin capsules

Brian Jones

CHAPTER CONTENTS

INTRODUCTION

The word capsule is derived from the Latin *capsula*, meaning a small box. In current English usage it is applied to many different objects, ranging from flowers to spacecraft. In pharmacy the word is used to describe an edible package made from gelatin or other suitable material which is filled with medicines to produce a unit dosage, mainly for oral use. There are two types of capsule, 'hard' and 'soft'; better adjectives would be 'two-piece' instead of hard, and 'one-piece' instead of soft. The hard capsule consists of two pieces in the form of cylinders closed at one end: the shorter piece, called the 'cap', fits over the open end of the longer piece, called the 'body'.

RAW MATERIALS

The raw materials used in the manufacture of both types of capsule are similar. Both contain gelatin, water, colourants and optional materials such as process aids and preservatives; in addition, soft capsules contain various plasticizers. The major pharmacopœiæ (European, Japanese and US) permit the use of gelatin or other suitable material, and in recent years hard capsules have been manufactured also from hydroxypropyl methylcellulose in order to produce a shell with a low moisture content (Ogura and Matsuura 1998).

Gelatin

Gelatin is the major component of the capsule and has been the material from which they have traditionally been made. The reason for this is that gelatin possesses five basic properties:

1. It is non-toxic, widely used in foodstuffs, and acceptable for use worldwide.

2. It is readily soluble in biological fluids at body temperature.
3. It is a good film-forming material, producing a strong flexible film. The wall thickness of a hard gelatin capsule is about 100 μm.
4. Solutions of high concentration, 40% w/v, are mobile at 50°C. Other biological polymers, such as agar, are not.
5. A solution in water or in a water–plasticizer blend undergoes a reversible change from a sol to a gel at temperatures only a few degrees above ambient. This is in contrast to other films formed on dosage forms, where either volatile solvents or large quantities of heat are required to cause this change of state, e.g. tablet film coating. These films are formed by spraying and have a structure that could be described as formed of overlapping plates, whereas the gelatin films are homogenous in structure, which gives them their strength.

Gelatin is a substance of natural origin that does not occur as such in nature. It is prepared by the hydrolysis of collagen, which is the main protein constituent of connective tissues (Jones 1987). Animal skins and bones are the raw materials used for the manufacture. There are two main types of gelatin: type A, which is produced by acid hydrolysis, and type B, which is produced by basic hydrolysis. The acid process takes about 7–10 days and is used mainly for animal skins, because they require less pretreatment than do bones. The basic process takes about 10 times as long and is used mainly for bovine bones. The bones must first be decalcified by washing in acid to give a soft sponge-like material, called ossein; calcium phosphates are produced as a byproduct. The ossein is then soaked in lime pits for several weeks.

After hydrolysis the gelatin is extracted from the treated material using hot water. The first extracts contain the gelatin with the highest physical properties, and as the temperature is raised the quality falls. The resulting weak solution of gelatin is concentrated in a series of evaporators and then chilled to form a gel. This gel is than extruded to form strands, which are then dried in a fluidized-bed system. The dried material is graded and then blended to meet the various specifications required. The properties of gelatin that are most important to the capsule manufacturers are the Bloom strength and the viscosity. The Bloom strength is a measure of gel rigidity. It is determined by preparing a standard gel (6.66% w/v) and maturing it at 10°C. It is defined as the load in grams required to push a standard plunger 4 mm into the gel. The gelatin used in hard capsule manufacture is of a higher bloom strength (200–250 g) than that used for soft capsules (150 g) because a more rigid film is required for the manufacturing process.

Many materials used in the manufacture of pharmaceuticals are manufactured from raw materials of bovine origin, e.g. stearates and gelatin. The outbreak of BSE, which started in the UK, has led to strict rules being introduced by the EU to minimize the risk of transmitting animal spongiform encephalopathy agents (TSEs). All parts of the bovine animals have been rated for infectivity and the high-risk parts, such as the brain and spinal cord, are removed before any processing takes place. The EU rules specify that the animals used must come from herds free from BSE, be subjected to pre- and postmortem veterinary inspection, and be processed by defined manufacturing processes by quality assured companies. No animal material from the UK, Eire, Switzerland or Portugal is permitted to be used. Pharmaceuticals and medicines are controlled specifically through a guidance document (CPMP/BWP/1230/98) issued by the European Agency for the Evaluation of Medicines Products (EMEA).

Colourants

The colourants that can be used are of two types: water-soluble dyes or insoluble pigments. To make a range of colours dyes and pigments are mixed together as solutions or suspensions. The dyes used are mostly synthetic in origin and can be subdivided in the azo dyes – those that have an –N=N– linkage – and the non-azo dyes, which come from a variety of chemical classes. Most dyes used currently are of the non-azo class, and the three most widely used are erythrosine (E127), indigo carmine (E132) and quinoline yellow (E104). Two types of pigment are used: iron oxides (E172), black, red and yellow, and titanium dioxide (E171), which is white and used to make the capsule opaque. The colourants that can be used to colour medicines are governed by legislation, which varies from country to country despite the fact that it is based on toxicological testing (Jones 1993). In the last few years there has been a move away from soluble dyes to pigments, particularly the iron oxides, because they are not absorbed on ingestion.

Process aids

The USNF describes the use of gelatin containing not more than 0.15% w/w of sodium lauryl sulphate for use in hard gelatin capsule manufacture. This

functions as a wetting agent, to ensure that the lubricated metal moulds are uniformly covered when dipped into the gelatin solution.

Preservatives were formerly added to hard capsules as an in-process aid in order to prevent microbiological contamination during manufacture. Manufacturers operating their plants to GMP guidelines no longer use them. In the finished capsules the moisture levels, 13.0–16.0% w/v, are such that they will not support bacterial growth because the moisture is too strongly bonded to the gelatin molecule.

MANUFACTURE

The process in use today is the same as that described in the original patent of 1846 (Jones 2000). Metal moulds at room temperature are dipped into a hot gelatin solution which gels to form a film. This is dried, cut to length, removed from the moulds and the two parts are joined together. The difference today is that the operation is now fully automated, carried out as a continuous process on large machines housed in air-conditioned buildings. There are only a comparatively few specialist companies that manufacture empty capsule shells for supply to the pharmaceutical and health-food industries, who fill them with their own products. Two companies, which have done most of the pioneering work in the field, have been making capsules for 100 years: Shionogi Qualicaps (formerly Eli Lilly & Co.) since 1897, and Warner Lambert's Capsugel (formerly Parke Davis) since 1902.

The first step in the process is the preparation of the raw materials. A concentrated solution of gelatin, 35–40%, is prepared using demineralized hot water, 60–70°C, in jacketed pressure vessels. This is stirred until the gelatin has dissolved and then a vacuum is applied to remove any entrapped air bubbles. Aliquots of this solution are then dispensed into suitable containers and the required amounts of dye solutions and pigment suspensions added. The viscosity is measured and adjusted to a target value by the addition of hot water. This latter parameter is used to control the thickness of the capsule shells during production: the higher the viscosity the thicker the shell wall produced. The prepared mixes are then transferred to a heated holding hopper on the manufacturing machine.

The manufacturing machines are approximately 10 m long, 2 m wide and 3 m high. They consist of two parts, which are mirror images of each other: on one half the capsule cap is made and on the other the capsule body. The machines are also divided into two levels, an upper and a lower. The moulds, commonly referred to as 'pins', are made of stainless steel and are mounted in sets on metal strips, called 'bars'. There are approximately 40 000 mould pins per machine. The machines are housed in large rooms where the humidity and temperature are closely controlled.

The sequence of events in the manufacturing process is shown in Figure 29.1. At the front end of the machine is a hopper, called a 'dip pan' or 'pot'. This holds a fixed quantity of gelatin at a constant temperature, between 45° and 55°C. The level of solution is maintained automatically by a feed from the holding hopper. Capsules are formed by dipping sets of moulds, which are at room temperature, 22°C, into this solution. A film is formed on the surface of each mould by gelling. The moulds are slowly withdrawn from the solution and then rotated during their transfer to the upper level of the machine, in order to form a film of uniform thickness. Groups of 'pin bars' are then passed through a series of drying kilns, in which large volumes of controlled humidity air are blown over them. When they reach the rear of the machine the bars are transferred back to the lower level and pass through further drying kilns until they reach the front of the machine. Here the dried films are removed from the moulds, cut to the correct length, the two parts joined together and the complete capsule delivered from the machine. The mould pins are then cleaned and lubricated for the start of the next cycle.

The machines are normally operated on a 24-hour basis 7 days per week, stopping only for maintenance. The output per machine is about 1 million capsules per day, depending upon the size: the smaller the capsule the higher the output.

The assembled capsules are not fully closed at this stage and are in a 'prelocked' position, which prevents them falling apart before they reach the filling machine. The capsules now pass through a series of sorting and checking processes, which can be either manual, mechanical or electronic, to remove as many defective ones as possible. The quality levels are checked through the process using standard statistical sampling plans based on the Military Inspection Standards. If required, capsules can be printed at this stage. This is done using an offset gravure roll printing process using an edible ink based on shellac. The information printed is typically either the product name or strength, a company name or logo, or an identification code. The capsules are finally packed for shipment in moisture-proof liners, preferably heat-sealed aluminium foil bags, in cardboard cartons. In these containers they can be

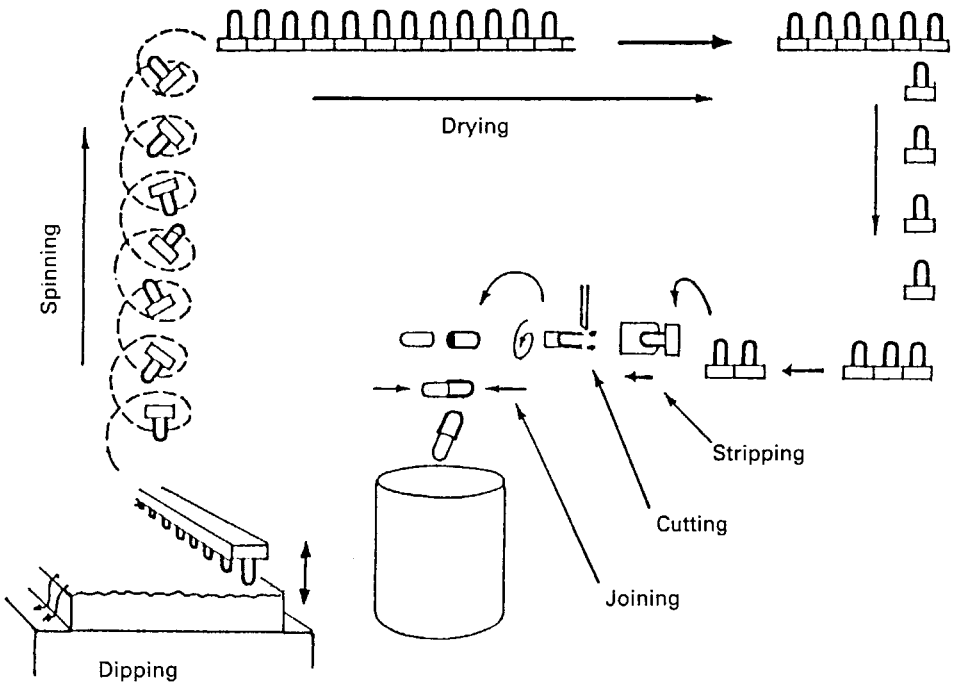


Fig. 29.1 The sequence of two-piece hard gelatin capsule shell manufacture.

stored for long periods without deterioration in quality, provided they are not subjected to localized heating or sudden temperature changes that will affect their moisture content and dimensions.

Empty capsule properties

Empty capsules contain a significant amount of water that acts as a plasticizer for the gelatin film and is essential for their function (Jones 2000). During industrial filling and packaging operations they are subjected to mechanical handling, and because the gelatin walls can flex these forces can be absorbed without any adverse effect. The standard moisture content specification for hard gelatin capsules is between 13.0% and 16.0% w/w. This value can vary depending upon the conditions to which they are exposed: at low humidities they will lose moisture and become brittle, and at high humidities they will gain moisture and soften. The moisture content can be maintained within the correct specification by storing them in sealed containers at an even temperature.

Capsules are readily soluble in water at 37°C. When the temperature falls below this their rate of solubility decreases. At below about 30°C they are insoluble and simply absorb water, swell and distort. This is an important factor to take into account during disintegration and dissolution testing. Because of this most Pharmacopœiae have set a limit of 37 ° ± 1°C for the media for carrying out these tests. Capsules made from HPMC have a different solubility profile, being soluble at temperatures as low as 10°C (Chiwele et al 2000).

Capsule filling

Capsule sizes

Hard gelatin capsules are made in a range of fixed sizes; the standard industrial sizes in use today for human medicines are from 0 to 4 (Table 29.1). For a powder the simplest way in which to estimate the fill weight is to multiply the body volume by its tapped bulk density (Jones 1998). For liquids, the fill weight is calculated by multiplying the specific

Table 29.1 Capsule size and body fill volumes

Capsule size	Body volume (mL)
0	0.67
1	0.48
2	0.37
3	0.28
4	0.20

gravity of the liquid by the capsule body volume × 0.8.

To accommodate special needs some intermediate sizes are produced, termed 'elongated sizes', that typically have an extra 10% of fill volume over the standard sizes, e.g. for 500 mg doses of antibiotics elongated size 0 capsules are commonly used. The shape of the capsule has remained virtually unchanged since its invention over 150 years ago, except for the development of the self-locking capsule. These were introduced during the 1960s, when automatic filling and packaging machines were introduced. Filled capsules were subjected to vibration during this process, causing some to come apart and spill their contents. To overcome this, modern capsule shells have a series of indentations on the inside of the cap and on the external surface of the body which, when the capsule is closed after filling, form an interference fit sufficient to hold them together during mechanical handling. The manufacturer of the empty shells can be identified from the types of indent, which are specific to each one.

Capsule shell filling

Hard gelatin capsules can be filled with a large variety of materials of different physicochemical properties. The limitations in the types of material that can be filled are shown in Table 29.2. Gelatin is a relatively inert material. The substances to be avoided are those which are known to react with it, e.g. formaldehyde, which causes a crosslinking reaction that makes the capsule insoluble, or those that interfere with the integrity of the shell, e.g. substances containing free water, which can be absorbed by the gelatin causing it to soften and distort.

The materials that have been filled into hard gelatin capsules are given in Table 29.3. The reason that such a range of materials can be handled is the nature of the capsule-filling process. Empty hard gelatin capsules are supplied in bulk containers. First, it is necessary for the filling machine to orientate them so that they are all pointing in the same direction, i.e. body first. To do this they are loaded into a hopper and from there pass down through tubes to a rectification section. Here the capsules are held in tight-fitting slots. Metal fingers strike them in the middle, and because the bodies have the smaller diameter, they rotate away from the direction of impact. Next the capsules are sucked through bushings that trap the caps, because of their greater diameter, separating them from the bodies. The bodies are then passed under the dosing mechanism and filled with material. Thus providing a substance can be measured and dosed, it can be filled into capsules. The caps are then repositioned over the bodies and metal fingers push the bodies up into them to rejoin the two parts.

Table 29.2 Limitations in properties of materials for filling into capsules

Must not react with gelatin
Must not contain a high level of 'free' moisture
The volume of the unit dose must not exceed the sizes of capsule available

Table 29.3 Types of material for filling into hard gelatin capsules

Dry solids
Powders
Pellets
Granules
Tablets
Semisolids
Thermosoftening mixtures
Thixotropic mixtures
Pastes
Liquids
Non-aqueous liquids

Capsule-filling machines

The same set of basic operations is carried out whether capsules are being filled on the bench for extemporaneous dispensing or on high-speed automatic machines for industrial products. The major difference between the many methods available is the way in which the dose of material is measured into the capsule body.

Filling of powder formulations

Bench-scale filling

There is a requirement for filling small quantities of capsules, from 50 to 10 000, in community pharmacy, in hospital pharmacy, or in industry for special prescriptions or trials. There are several simple pieces of equipment available for doing this, e.g. the 'Feton'

from Belgium or the 'Labocaps' from Denmark. These consist of sets of plastic plates which have predrilled holes to take from 30 to 100 capsules of a specific size. Empty capsules are fed into the holes, either manually or with a simple loading device. The bodies are locked in their plate by means of a screw and the caps in their plate are removed. Powder is placed on to the surface of the body plate and is spread with a spatula so that it is filled into the bodies. The uniformity of fill weight is very dependent upon good flow properties of the powder. The cap plate is then repositioned over the body one and the capsules are rejoined using manual pressure.

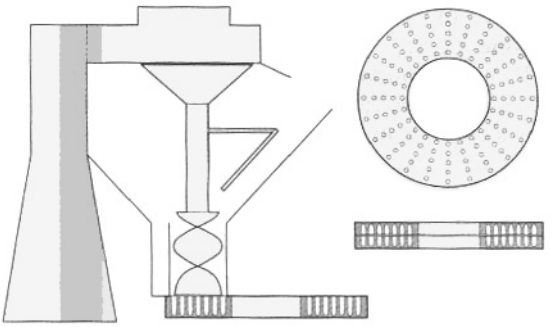


Fig. 29.2 An auger filling machine using the ring system, Model No. 8 (from Jones 2001, with permission).

Industrial-scale filling

The machines for the industrial-scale filling of hard gelatin capsules come in great variety of shapes and sizes, varying from semi- to fully automatic and ranging in output from 5000 to 15 000 per hour. Automatic machines can be either continuous in motion, like a rotary tablet press, or intermittent, where the machine stops to perform a function and then indexes round to the next position to repeat the operation on a further set of capsules.

The dosing systems can be divided into two groups:

- *Dependent*: dosing systems that use the capsule body directly to measure the powder. Uniformity of fill weight can only be achieved if the capsule is filled completely.
- *Independent*: dosing systems where the powder is measured independently of the body in a special measuring device. Weight uniformity is not dependent on filling the body completely. With this system the capsule can be part filled.

Dependent dosing systems

The auger Empty capsules are fed into a pair of ring holders (Fig. 29.2), the caps being retained in one half and the bodies in the other. The body holder is placed on a variable-speed revolving turntable; the powder hopper is pulled over the top of this plate, which revolves underneath it. In the hopper a revolving auger forces powder down into the capsule bodies. The weight of powder filled into the body is dependent mainly upon the time the body is underneath the hopper during the revolution of the plate holder.

These machines are semiautomatic in operation, requiring an operator to transfer the capsule holders from one operation to the next. It was the first developed for large-scale use at the beginning of the 20th century and is still widely used in many countries. The contact parts of these machines were originally made from cast iron, but are now made from stainless steel to comply with GMP requirements. The machines are manufactured by many different makers, but are all based on the original Colton Model No. 8 design. Their output varies between 15 000 and 25 000 per hour and is dependent upon the skill of the operator.

Independent dosing systems Most industrial machines in use in Europe and the USA are fully automatic and use dosing mechanisms that form a 'plug' of powder. This is a soft compact formed at low compression forces – between 10 and 100 N – which are significantly less than those used in tabletting. The reason the plug is soft is because it is not the final dosage form, unlike the tablet, as the material will be contained inside a capsule shell. There are two types of plug-forming machine: those that use a 'dosator' system and those that use a 'tamping finger and dosing disc' system.

Dosators This consists of a dosing tube inside which there is a movable spring-loaded piston, thus forming a variable-volume chamber in the bottom of the cylinder (Fig. 29.3). The tube is lowered open end first into a bed of powder, which enters the tube to fill the chamber and form a plug. This can be further consolidated by applying a compression force with the piston. The assembly is then raised from the powder bed and positioned over the capsule body. The piston is lowered, ejecting the powder plug into the capsule body. The weight of powder filled can be adjusted by altering the position of the piston inside the tube, i.e. increasing or decreasing the volume, and by changing the depth of the powder bed.

This system is probably the most widely used in the world and is the one that is described the most in the literature. Examples of machines that use this system are:

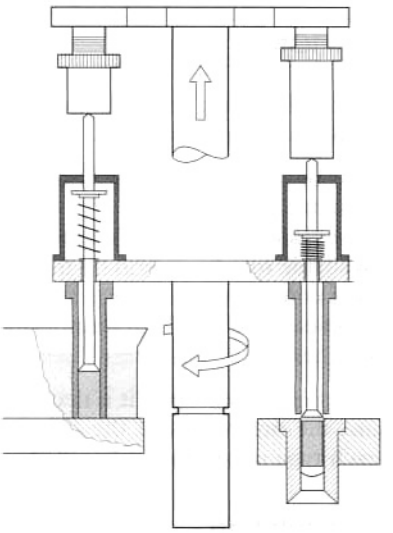


Fig. 29.3 A dosing tube or dosator-type machine, Zanasi RM63 (from Jones 2001, with permission).

- *Intermittent motion*: Zanasi (IMA), Pedini, Macophar and Bonapace. Their outputs range from 5000 to 60 000 per hour.
- *Continuous motion*: MG2, Matic (IMA). Their outputs range from 30 000 to 150 000 per hour.

Tamping finger and dosing disc The dosing disc forms the bottom of a revolving powder hopper (Fig. 29.4). This disc has in it a series of sets of accurately drilled holes in which powder plugs are formed by several sets of tamping fingers – stainless steel rods that are lowered into them through the bed of powder. At each position the fingers push material into the holes, building up a plug before they index on to the next position. At the last position the finger pushes the plug through the disc into a capsule body. The powder fill weight can be varied by the amount of insertion of the fingers into the disc, by changing the thickness of the dosing disc, and by adjusting the amount of powder in the hopper.

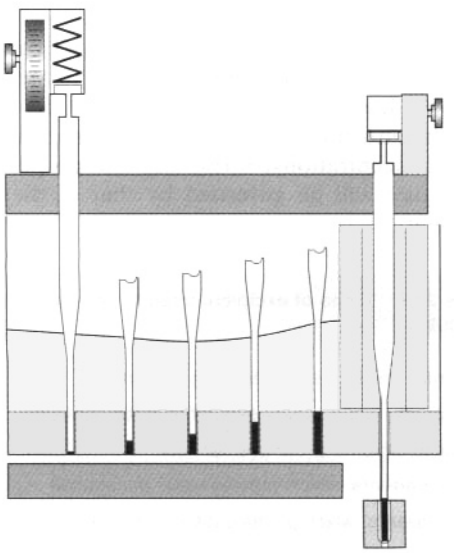


Fig. 29.4 A dosing disc and tamping finger machine, Höfliger & Karg (from Jones 2000, with permission).

The machines that use this system are all intermittent in motion. Examples are the Höfliger and Karg, manufactured by Robert Bosch, and the Shionogi Qualicaps F-80.

Instrumented capsule-filling machines and simulators

Unlike tablet machines, few workers have instrumented capsule-filling machines. This is for a variety of reasons. Capsules are used only in the pharmaceutical and health-food industries, as opposed to tablets, which are widely used by many other industries and for which there is therefore more incentive to do fundamental research. The tablet press is simple to quantify: there are two punches and a die that holds a specific volume of material. On a capsule-filling machine there are a variety of moving parts involved in dosing, which occurs in an unconfined bed of powder. The forces involved are small. As a result of this, comparatively few papers have been published on the topic. Machines that use the dosator system have been studied the most (Augsburger 1988). Strain gauges have been fixed to the piston that have enabled the compression forces – 10–250 N – and ejection forces – 1–20 N – in lubricated products to be measured. Distance transducers have been used to measure the relative movements of the piston and dosator. Simulators have also been built to overcome the problem of the machine parts moving, but to date these have had limited application (Britten et al 1995).

Pellet filling

Preparations formulated to give modified-release patterns are often produced as granules or coated pellets. They are filled on an industrial scale using machines adapted from powder use. All have a dosing system based on a chamber with a volume that can easily be changed. Pellets are not compressed in the process and may have to be held inside the measuring devices by mechanical means, e.g. either by inverting the dosator or by applying a

suction to the dosing tube. In calculating the weight of particles that can be filled into a capsule it is necessary to make an allowance for their size. Unlike powders, which have a much smaller size, they cannot fill all the available space within the capsule because of packing restrictions. The degree of this effect will be greater the smaller the capsule size and the larger the particle diameter.

Tablet filling

Tablets are placed in hoppers and allowed to fall down tubes, at the bottom of which is a gate device that will allow a set number of tablets to pass. These fall by gravity into the capsule bodies as they pass underneath the hopper. Most machines have a mechanical probe that is inserted into the capsule to check that the correct number of tablets has been transferred. Tablets for capsule filling are normally film coated to prevent dust, and are sized so that they can fall freely into the capsule body.

Semisolid and liquid filling

Liquids can easily be dosed into capsules using volumetric pumps (Jones 2001). The problem after filling is to stop leakage from the closed capsule. This can be done in one of two ways, either by formulation or by sealing of the capsule. Semisolid mixtures are formulations that are solid at ambient temperatures and can be liquefied for filling by either heating thermosoftening mixtures, or by stirring thixotropic mixtures. After filling they cool and solidify, or revert to their resting state in the capsule to form a solid plug. Both types of formulations are filled as liquids using volumetric pumps. These formulation are similar to those that are filled into soft gelatin capsules, but differ in one important respect: they can have melting points higher than 35°C, which is the maximum for soft gelatin capsules because this is the temperature used by the sealing rollers during their manufacture. Non-aqueous liquids, which are mobile at ambient temperatures, require the capsules to be sealed after filling. The industrially accepted method for this is to seal the cap and body together by applying a gelatin solution around the centre of the capsule after it has been filled. When this has been dried it forms a hermetic seal that prevents liquid leakage, contains odours inside the shell and significant reduces oxygen permeation into the contents, protecting them from oxidation. An example of such equipment is the Shionogi Qualicaps Hicapseal machines, which have outputs ranging from 10 000 to 100 000 per hour.

FORMULATION

All formulations for filling into capsules have to meet the same basic requirements:

1. They must be capable of being filled uniformly to give a stable product.
2. They must release their active contents in a form that is available for absorption by the patient.
3. They must comply with the requirements of the Pharmacopœiæ and regulatory authorities, e.g. dissolution tests.

In order to formulate rationally it is necessary to take into account the mechanics of the filling machines and how each type of product is handled.

Powder formulation

The majority of products for filling into capsules are formulated as powders. These are typically mixtures of the active ingredient together with a combination of different types of excipients (Jones 1995; Table 29.4). The ones selected depend upon several factors:

- The properties of the active drug
- Its dose, solubility, particle size and shape
- The size of capsule to be used.

The latter factor defines the free space inside the capsule that is available to the formulator (Jones 1998). The easier active compounds to formulate are low-dose potent ones, which in the final formulation occupy only a small percentage of the total volume – <20% – and so the properties of the mixture will be governed by the excipients chosen, whereas those compounds with a high unit dose, e.g. 500 mg of an antibiotic, leave little free space within the capsule and excipients must be chosen that exert their effect at low concentrations, <5%, and the properties of the mixture will be governed by that of the active ingredient.

Table 29.4 Types of excipient used in powder-filled capsules

Diluents, which give plug-forming properties
Lubricants, which reduce powder to metal adhesion
Glidants, which improve powder flow
Wetting agents, which improve water penetration
Disintegrants, which produce disruption of the powder mass
Stabilizers, which improve product stability

Formulation for filling properties

There are three main factors in powder formulation:

- Good flow, (using free-flowing diluent and glidant)
- No adhesion (using lubricant)
- Cohesion (plug-forming diluent).

The factor that contributes most to the uniform filling of capsules is good powder flow. This is because the powder bed, from which the dose is measured, needs to be homogeneous and packed reproducibly in order to achieve uniform fill weights. Packing is assisted by mechanical devices on the filling machines. Low-dose actives can be made to flow well by mixing them with free-flowing diluents, e.g. lactose. The diluent is chosen also for its plug-forming properties: the most frequently used ones are lactose, maize starch and microcrystalline cellulose. When space is limited then either glidants, which are materials that reduce interparticulate friction, such as colloidal anhydrous silica, or lubricants, which are materials that reduce powder to metal adhesion, e.g. magnesium stearate, are added, enabling the dosing devices to function efficiently. Both of these types of material exert their effect by coating the surfaces of the other ingredients, and thus the mixing of these into the bulk powder has a significant effect on their functioning.

Formulation for release of active ingredients

The first stage in active ingredient release is disintegration of the capsule shell. When capsules are placed in a suitable liquid at body temperature, (37°C) the gelatin starts to dissolve and within 1 minute the shell will split, usually at the ends. With a properly formulated product the contents will start to empty out before all the gelatin has dissolved. The official tests for disintegration and dissolution were originally designed for tablets. Capsules have very different physical properties, and after the contents have emptied out the gelatin pieces remaining will adhere strongly to metal surfaces and may confuse the end-point of the test.

The literature shows that the rate-controlling step in capsule disintegration and product release is the formulation of the contents, which ideally should be hydrophilic and dispersible (Jones 1987). The factors that can be modified to make the active ingredients readily available depend upon their properties and those of any excipients being used. The active ingredients have a fixed set of physico-chemical properties which, except for the particle size, are out of the control of the formulator.

It has been shown that the particle size influences the rate of absorption for several compounds. For sulfisoxazole (Fig. 29.5) three different particle sizes were filled into capsules and administered to dogs; the smallest particle gave the highest peak blood level. This can be explained simply by the fact that the solution rate is directly proportional to the surface area of the particles: the smaller the particle the greater the relative surface area. However, this is not a panacea for formulation problems because small particles tend to aggregate together and the effect is lost. It has been shown that the important factor with particle size was the 'effective surface area', which is the area of the active available to the dissolution fluid. This is related to the packing of particles and is a measure of how well the fluid can penetrate into the mass.

Diluents are the excipients that are usually present in the greatest concentration in a formulation. They were classically defined as inert materials added to a mixture to increase its bulk to a more manageable quantity. Although they are relatively inert chemically, they do play a role in release. The case that first demonstrated this happened in Australia in the late 1960s. A capsule was reformulated that contained diphenylhydantoin, which is used for the treatment

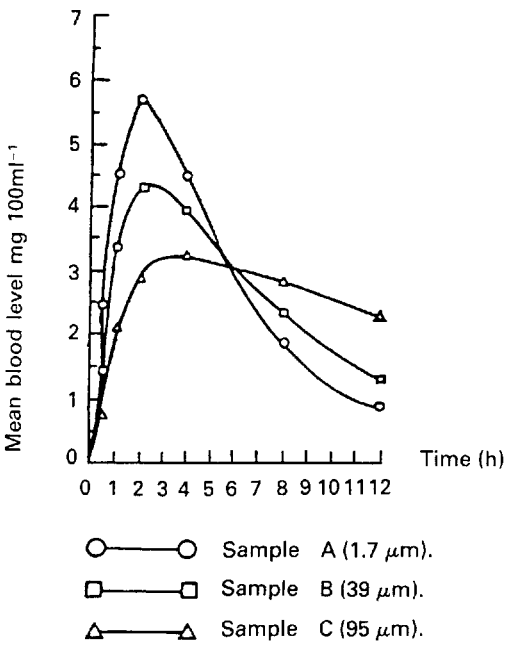


Fig. 29.5 Effect of particle size on bioavailability (after Fincher et al 1965, with permission).

of epilepsy and is taken chronically. The diluent used was changed from calcium sulfate to lactose. In the months following this change there was an upsurge in reports of side-effects similar to overdosing of product. It was demonstrated that the change had had a significant effect on the bioavailability of the active (Fig. 29.6). The change to lactose gave much higher blood levels of the drug, which was probably due to the fact that it is readily soluble whereas calcium sulfate is not.

Since this occurrence the phenomenon has been shown to occur with other actives. The diluent used should be chosen in relationship to the solubility of the active. If a soluble diluent such as lactose is added to a poorly or insoluble compound it will make the powder mass more hydrophilic, enabling it to break up more readily on capsule shell disintegration. The converse is also true: actives that are readily soluble are best mixed with insoluble diluents such as starch or microcrystalline cellulose, because they help the powder mass to break up without interfering with their solubility in the medium.

Some excipients, such as lubricants and glidants, are added to formulations to improve their filling properties, and these can sometimes have an effect on release. The important thing to avoid in formulations are materials that tend to make the mass more hydrophobic. The most commonly used lubricant for both encapsulation and tabletting is magnesium stearate. Simmons et al (1972) studied the dissolution rate of chlordiazepoxide formulations with three levels of magnesium stearate, 0%, 1% and 5% (Fig. 29.7). They found that the dissolution rate was greatly reduced at the highest level of magnesium stearate, which they explained was due to the poor wetting of the powder mass. However, hydrophobic additives are

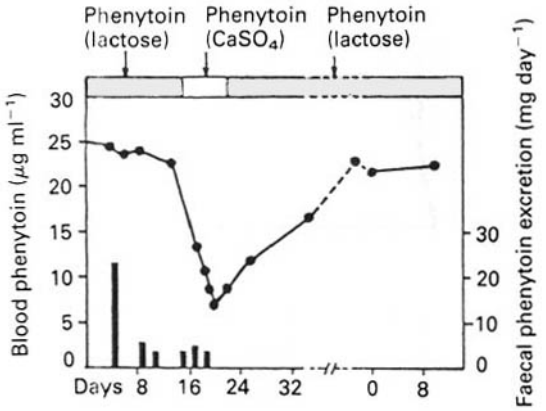


Fig. 29.6 Effect of diluent on bioavailability (after Tyrer et al 1970, with permission).

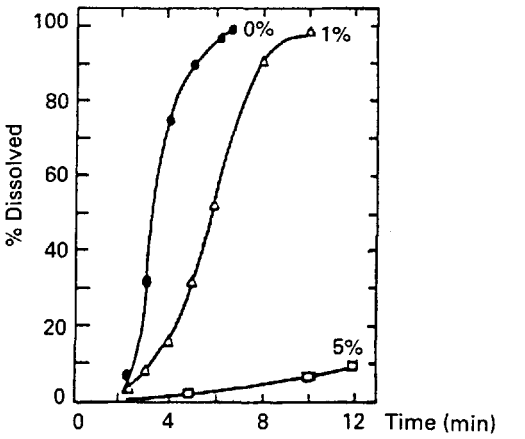


Fig. 29.7 Effect of lubricant on release of active ingredient (after Simmons et al 1972, with permission).

not always deleterious because they reduce the cohesiveness of the powder mass. This was first demonstrated by Nakagawa et al (1980), who were studying the dissolution of different particle sizes of rifampicin with and without magnesium stearate (Fig. 29.8). They found that for the larger particles (180–355 μm) the addition of magnesium stearate reduced the rate, but for the smaller particles (<75 μm) it increased the rate. This is because magnesium stearate reduces the cohesiveness of the small particles so that they spread more rapidly through the dissolution medium than the unlubricated material. Augsburger et al (1988) studied the system hydrochlorthiazide, microcrystalline cellulose and various levels of magnesium stearate (Fig. 29.9). They filled capsules on an instrumented machine using the same compression force, and found that as the concentration of magnesium stearate increased the dissolution rate improved to a maximum value at about 1.0% w/v, after which it fell. They correlated this to the hardness of the powder plug, which followed a similar pattern, becoming

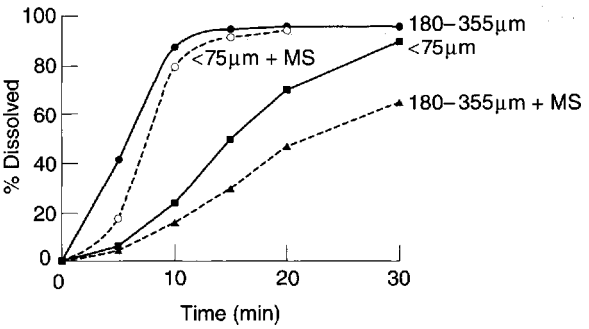


Fig. 29.8 Effect of lubricant on in vitro release of rifampicin (after Nakagama et al 1980, with permission).

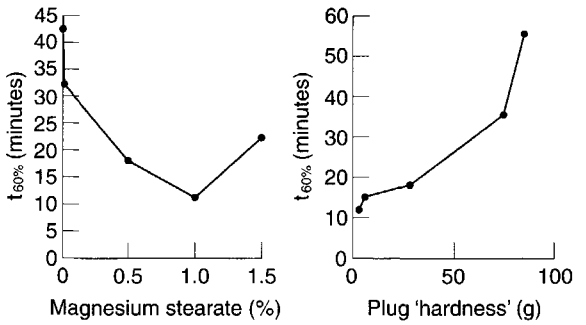


Fig. 29.9 Effect of lubricant on in vitro release of hydrochlorthiazide (after Botzolakis et al 1982, with permission).

softer – i.e. easier to break apart – as the concentration of lubricant increased. Above 1.0% the plug becomes too hydrophobic for the increase in 'softness' to compensate for this.

The increase in the use of dissolution testing for control purposes has led to products being formulated for this factor. This has been achieved in two ways: by the addition of either surfactants or disintegrants. The addition of a wetting agent, sodium lauryl sulphate, has been studied by several workers. For poorly soluble drugs the use of a soluble diluent together with 1% sodium lauryl sulphate gave the best results. Disintegrants were formerly never added to capsule formulations because starch, which was the most widely used tabletting disintegrant, does not function well in this context. This is because the powder plug is much less compacted than a tablet and the starch swells insufficiently to disrupt it. In more recent times '***superdisintegrants***' have been introduced that either swell many fold on absorbing water, e.g. sodium starch glycolate and croscarmellose, or that act as wicks, attracting water into the plug, e.g. crospovidone. These actions are sufficient to help break up the capsule plug. The choice of disintegrant is dependent upon the solubility of the active and the diluent, which governs whether either swelling or wicking is the main disruptive force required (Botzolakis and Augsburger 1988).

Formulation optimization

The formulator has to produce a product that complies with the three formulation goals. Sometimes these are contradictory: for example, extra hydrophobic lubricant is required for filling machine performance, which could interfere with release. Therefore, in the development stage the formulation needs to be optimized so that it can meet the product specification. This can be done by using various statistical tools based on analysis of variance experiments that can identify the contribution of each excipient and process operation to the product performance, e.g. uniformity of fill weight and content, dissolution rate, disintegration, yield etc. (Armstrong and James 1996).

The latest computer-based systems to aid the formulator are 'expert systems', which are based on neural networks coupled to a knowledge base (Lai et al 1996). The systems are set up using rules that have been established through experience and research. They enable a formulator to enter into the system the characteristics of the active ingredient and the type of product they would like to produce. The system then comes back with suggested formulations to try. This can significantly reduce the development time for a new product.

To summarize, the main factors in powder formulation release are:

- Active ingredient, optimum particle size
- Hydrophilic mass, relating solubility of active to excipients
- Dissolution aids, wetting agent, superdisintegrant
- Optimum formulation for filling and release.

Formulation for position of release

Many products are formulated to release their contents in the stomach. However, this may not always be the best place for the absorption of the active ingredient, and capsule formulation can be readily manipulated to release their contents at various positions along the gastrointestinal tract (Jones 1991).

A common problem with oral dosage forms is making them easy to swallow. Certain people have great difficulty with this because the process is not a reflex and is controlled by the central nervous system. The capsule is a good shape for swallowing because the tongue will automatically align it with its long axis pointing towards the throat. Many tablets are now made this shape – called a 'caplet' – in order to facilitate swallowing. Patients who have difficulty swallowing should be instructed to do it either standing or sitting, in order to make full use of gravity, and to take a drink of water to lubricate the throat. They should drink a little water and hold it in the mouth. The capsule should be placed in the mouth and the head tilted forward. The capsule will now float on the water towards the back of the mouth and when the head is lifted the bolus of water and the capsule will go straight down the throat to the stomach.

In the stomach the release of the active ingredient can be modified in a number of ways. It has been

suggested that for some compounds the best way to improve their absorption is for the dosage form to be retained in the stomach so that it will dissolve slowly, releasing a continuous flow of solution into the intestines. 'Floating capsules' have been made which contain various hydrophilic polymers, such as methylcellulose, that swell on contact with water and form a mass that can float on the gastric liquids. Some compounds are destroyed at acid pHs, and an enteric product can be made by either coating the filled capsule with an enteric film in a similar manner to a tablet, or by formulating the contents as pellets and coating them with an enteric polymer.

Much has been written in the literature about the advantages or disadvantages of making prolonged-release dosage forms as monolithic or as multi-particulate systems. The current consensus is that multiparticulates are better because they will be released in a stream from the stomach when the capsule shell disintegrates, and will not be retained for variable periods of time as would a monolithic product. They also avoid the risk of the dose being dumped at one point, which could cause problems of local gastric irritation.

Certain compounds are absorbed only at specific locations along the intestines. If this window of absorption is known then a formulation can be made to release its contents in that region. There is currently an interest in targeting compounds to the distal parts of the intestines. This has been achieved in two ways. Products can be formulated to give a prolonged release and filled into a capsule that is enteric coated, e.g. Colpermin (Pharmacia Upjohn), an enteric-coated capsule filled with a prolonged-release formulation of peppermint oil. The capsule disintegrates in the duodenum and the contents slowly release the peppermint oil, which acts as a smooth muscle relaxant as it passes through the remainder of the tract. Products have also been prepared that have been coated with polymers that are enteric and are soluble only at higher pHs, 6–7. This pH is not reached until further along the small intestine, and so the contents are delivered to the more distal parts. Currently many new chemical entities are proteins or polypeptides, and to make an oral dosage form it is necessary to deliver them to the colon, thereby avoiding the proteolytic enzymes in the stomach and small intestine. The release mechanisms for these capsules are based on specific colonic conditions, e.g. coatings that are disrupted by colon-specific enzymes or by pressure (Nagata and Jones 2001).

Not all capsules administered by the oral route are intended for the gastrointestinal tract. Capsules have been used successfully for many years for products for inhalation. The active ingredient, which is micronized, is filled into the capsule either 'as is' or dispersed on a carrier particle. The weight filled into a capsule is much lower than for other types of product, typically less than 25 mg. These formulations are filled on automatic machines that have microdosing devices, and the product is administered by using a special inhaler. A capsule is placed in the device and the powder is released either by the halves of the capsule being forced apart or by the capsule wall being punctured by sharp pins. The device is breath actuated. When the patient breathes in a turbulent airflow dislodges the carrier particles (if present) and the active powder is inhaled into the lungs. The system has the added advantage over an aerosol device in that the patient can see how many doses they have taken by counting the number of capsules remaining.

REFERENCES

Armstrong, N.A. and James, K.C. (1996) *Pharmaceutical experimental design and interpretation*. Taylor & Francis, London.

Augsburger, L.L. (1988) S.T.P. *Pharma*, **4**, 116–122.

Botzolakis, J.E. and Augsburger, L.L. (1988) *J. Pharm. Pharmacol.*, **36**, 77–84.

Britten, J.R., Barnett, M.I., Armstrong, N.A. (1995) *Pharm. Res.*, **12**, 196–200.

Chiwele, I., Jones, B.E., Podczeck, F., *Chem. Pharm. Bull.*, 2000 (in press).

Fincher, J.H., Adams, H.M. and Beal, H.M. (1965) *J. Pharm. Sci.*, **54**, 704–708.

Jones, B.E. (1987) S.T.P. *Pharma*, **3**, 777–783.

Jones, B.E. (1991) S.T.P. *Pharma Pratiques*, **2**, 128–134.

Jones, B.E. (1993) *Pharm. Tech. Int.* **5(4)**, 14–20.

Jones, B.E. (1995) *Pharm. Tech. Europe*, **7(10)**, 25, 28–30, 34.

Jones, B.E. (1998) S.T.P. *Pharma Sci.*, **8(5)**, 277–283.

Jones, B.E. (2001) Hard capsules. In: Swarbrick, J. and Boylan, J.C. (eds) *Encyclopedia of Pharmaceutical Technology*, 2nd edn. Vol. 1 Marcel Dekker, New York.

Jones, R.T. (1987) Gelatin structure and manufacture and Gelatin physical and chemical properties. In: Ridgway K. (ed) *Hard Capsules, Development and Technology*. Pharmaceutical Press, London, Chs 2 & 3.

Lai, S., Podczeck, F., Newton, J.M., Daumesnil, R. (1996) *Pharm. Tech. Europe*, **8(9)**, 60, 62, 64, 66, 68.

Nagata, S. and Jones, B.E. (2001) *Eur. Pharm. Rev.* **5(2)**, 41–46.

Nakagawa, H., Mohri, K., Nakashima, K., Sugimoto, I. (1980) *Yakugaku Zasshi*, **100**, 1111–1117.

Ogura, T. and Matsuura, S. (1998) *Pharm. Tech. Europe*, **10(11)**, 32, 34, 36, 40, 42.

Simmons, D.L., Frechette, M., Ranz, R.J. Chen, W.S. and Patel, N.K. (1972) *Can. J. Pharm. Sci.*, 7, 62–65.

Tyrer, J.M., Eacke, M.J., Sutherland, J.M., Hopper, W.D. (1970) *Br. Med. J.*, **IV**, 271–273.

30
Soft gelatin capsules

Keith Hutchison, Josephine Ferdinando

CHAPTER CONTENTS

INTRODUCTION

The term 'soft gelatin capsules' is commonly abbreviated to 'softgels', and although the terms are interchangeable 'softgels' will be used throughout this chapter for the sake of convenience and consistency.

When pharmaceutical formulation scientists are faced with the challenge of designing a solid oral dosage form for drug compounds, they have a number of choices. Over recent years, new drug molecules have tended to be more hydrophobic and therefore less soluble in aqueous systems. In the case of drugs for oral administration, it is becoming more difficult to formulate poorly water-soluble drugs into products from which the drug is fully released and well absorbed. One of the best methods to overcome this problem is to make a liquid formulation containing the drug. In order to convert this liquid formula into a solid dosage form, it may be encapsulated into soft gelatin capsules.

This chapter explains the reasons why soft gelatin capsules are selected for formulation development, how they are formulated, and how they are manufactured.

DESCRIPTION OF THE SOFT GELATIN CAPSULE DOSAGE FORM (SOFTGELS)

Softgels consist of a liquid or semisolid matrix inside a one-piece outer gelatin shell (Fig. 30.1). Ingredients that are solid at room temperature can also be encapsulated into softgels, provided they are at least semisolid below approximately 45°C. The drug compound itself may be either in solution or in suspension in the capsule-fill matrix. The characteristics of the fill matrix may be hydrophilic (for example polyethylene glycols) or lipophilic (such as triglyceride vegetable oils). Indeed, in many formulations, the

matrix may be a mixture of both hydrophilic and lipophilic ingredients.

Significant advances have been made in recent years in the formulation of softgel fill matrices (Fig. 30.1). These include microemulsions and nanoemulsions encapsulated as preconcentrates in softgels. The term 'preconcentrate' means that the softgel fill matrix is a combination of lipophilic and hydrophilic liquids as well as surfactant components, which after oral administration disperse to form, for example, a microemulsion. If the dispersion results in even smaller droplets in the nanoparticle range, then the dispersion is known as a nanoemulsion.

The softgel capsule shell consists of gelatin, water and a plasticizer. It may be transparent or opaque, and can be coloured and flavoured if desired. Preservatives are not required owing to the low water activity in the finished product. The softgel can be coated with enteric-resistant or delayed-release material. Although virtually any shape softgel can be made, oval or oblong shapes are usually selected for oral administration.

Softgels can be formulated and manufactured to produce a number of different drug delivery systems:

- Orally administered softgels containing solutions or suspensions that release their contents in the stomach in an easy to swallow, convenient unit dose form (Fig. 30.2). This is the most common type of softgel, already described above;
- Chewable softgels, where a highly flavoured shell is chewed to release the drug liquid fill matrix. The drug(s) may be present in both the shell and the fill matrix;
- Suckable softgels, which consist of a gelatin shell containing the flavoured medicament to be sucked and a liquid matrix or just air inside the capsule;
- Twist-off softgels, which are designed with a tag to be twisted or snipped off, thereby allowing access to the fill material. This type of softgel can be very useful for unit dosing of topical

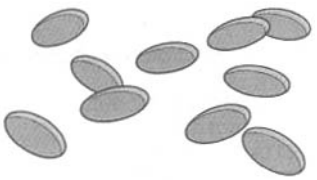


Fig. 30.2 Swallowable softgel capsules.

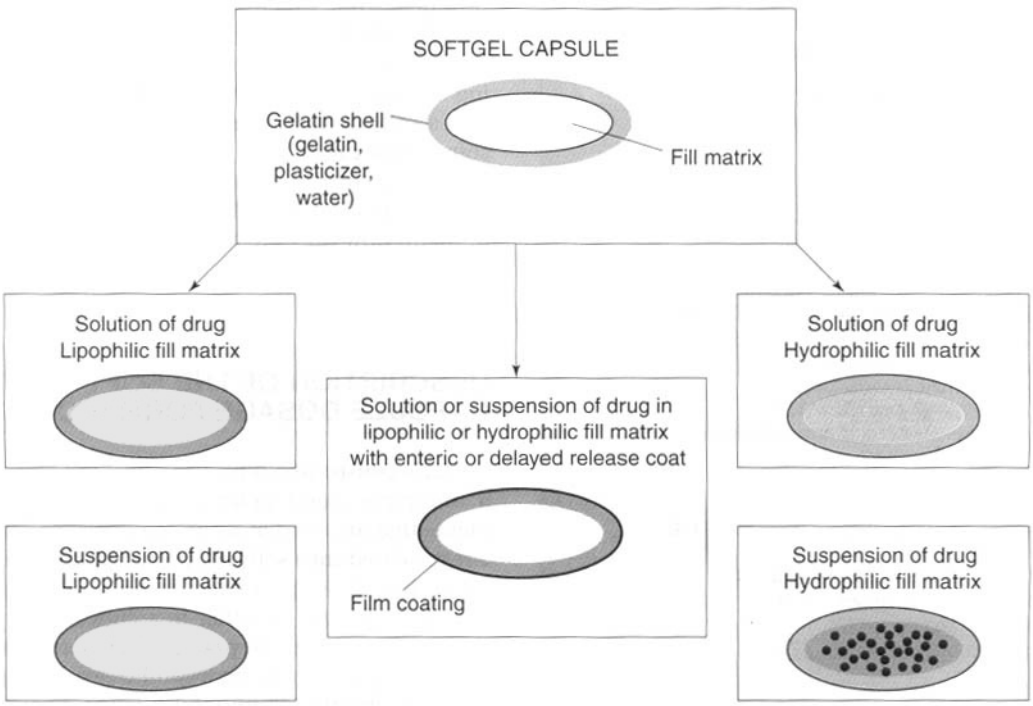


Fig. 30.1 Schematic diagram of different softgel formulations.

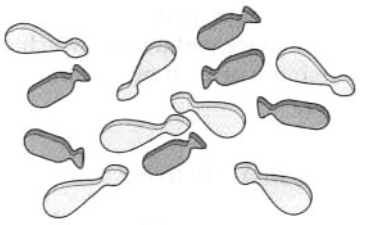


Fig. 30.3 Twist-off softgel capsules.

medication, inhalations, or indeed for oral dosing of a paediatric product (Fig. 30.3);
- Meltable softgels, designed for use as 'patient-friendly' pessaries or suppositories.

RATIONALE FOR THE SELECTION OF SOFTGELS AS A DOSAGE FORM

There are a number of reasons why softgels may be selected as the most suitable dosage form. These are summarized in Table 30.1. In the majority of cases improved drug absorption is the primary reason (Ferdinando 2000). However, the other features listed should also be remembered because, either individually or collectively, they are important factors that determine the selection of this drug delivery system.

Improved drug absorption

Increased rate of absorption

Major advances have been made in the development of softgel formulations to address particular drug absorption issues. One of the best methods is presentation of the drug to the gastrointestinal tract in the form of a solution from which it can be rapidly absorbed. This can be achieved using a drug–solution matrix in a softgel formulation whereby absorption is significantly faster than from other solid oral dosage forms, such as compressed tablets. This is because absorption of a poorly soluble drug from a tablet formulation is rate-limited by the need for disintegration into granules, then drug dissolution into gastrointestinal fluid. With the solution–softgel approach, the shell ruptures within minutes to release the drug solution, which is usually in a hydrophilic or highly dispersing vehicle that aids the rate of absorption. This may be a valuable attribute (a) for therapeutic reasons, such as the treatment of migraine or acute pain, or (b) where there is a limited absorptive region or 'absorption window' in the gastrointestinal tract. Figure 30.4 shows the faster absorption that can

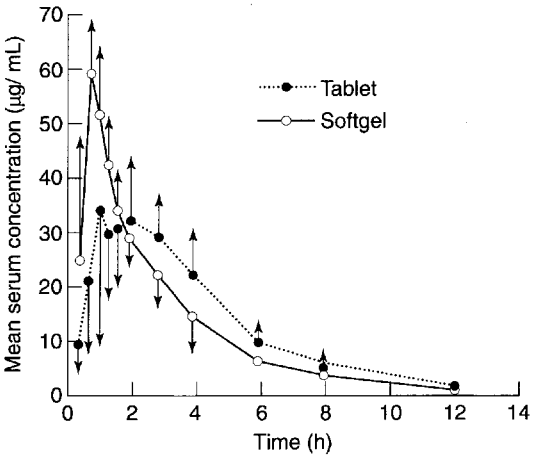


Fig. 30.4 Pharmacokinetic evaluation of softgels and tablets containing 400 mg ibuprofen (in 12 volunteers) (Saano 1991).

Table 30.1 Summary of the key features and advantages of the softgel dose form

Features	Advantages
Improved drug absorption	Improved rate and extent of absorption and/or reduced variability, mainly for poorly water-soluble drugs
Patient compliance and consumer preference	Easy to swallow. Absence of poor taste or other sensory problem. Convenient administration of a liquid-drug dosage form
Safety – potent and cytotoxic drugs	Avoids dust handling problems during dosage form manufacture: better operator safety and environmental controls
Oils and low melting-point drugs	Overcomes problems with manufacture as compressed tablet or hard-shell capsules
Dose uniformity for low-dose drugs	Liquid flow during dosage form manufacture is more precise than powder flow. Drug solutions provide improved homogeneity over powder or granule mixtures
Product stability	Drugs are protected against oxidative degradation by lipid vehicles and softgel capsule shells

be achieved using a solution–softgel formulation compared to a tablet.

Increased bioavailability

As well as increasing the rate of absorption, softgels may also improve the extent of absorption. This can be particularly effective for hydrophobic drugs with a relatively high molecular weight. An example of such a product is the protease inhibitor saquinavir, which has been formulated as a solution–softgel product (Perry and Noble 1988). The solution–softgel formulation provided around three times the bioavailability of saquinavir as measured by the area under the plasma–time curve (AUC), compared to a hard-shell capsule formulation.

In some cases a drug may be solubilized in a vehicle that is capable of spontaneously dispersing into an emulsion on contact with gastrointestinal fluid. This is known as a self-emulsifying system. In other cases a drug may be dissolved in an oil/surfactant vehicle that produces a microemulsion or a nanoemulsion on contact with gastrointestinal fluids. A nanoemulsion of progesterone has been developed that provides a good example of this type of formulation. The vehicle, consisting of oils and surfactants in appropriate proportions, when in contact with aqueous fluids, produces an emulsion with an average droplet size less than 100 nm. The solubility of the drug is maintained as long as possible, delivering solubilized drug directly to the enterocyte membrane. This produces increased bioavailability compared to formulations where the drug is dosed in the solid state. Figure 30.5 shows the plasma concentration–time profile for progesterone absorbed from the nanoemulsion formulation.

Softgel formulations may contain excipients, for example one or more surfactants which can aid the stability, wettability or even permeability of the drug (Aungst 2000).

Decreased plasma variability

High variability in drug plasma levels is a common characteristic of drugs with limited bioavailability. By dosing drug optimally in solution, the plasma level variability of such drugs can be significantly reduced. The cyclic polypeptide drug cyclosporine (Sandimmun Neoral®) benefits from such an approach by using a microemulsion preconcentrate in a softgel (Drewe et al 1992, Meinzer 2000).

Patient compliance and consumer preference

A number of self-medicating consumer preference studies have been carried out in an attempt to gauge the relative perception of softgels compared to hard-shell capsules and tablets. The results showed consistently that softgels were perceived to be appealing dosage forms to most consumers, and outperformed all other dosage forms in answering patient needs. Consumers expressed their preference for softgels in terms of (a) ease of swallowing, (b) absence of taste and (c) convenience.

One further aspect of improved compliance is that if, by using a drug solution in a softgel delivery system, its bioavailability is enhanced, it may be possible to reduce the dose administered in order to achieve therapeutic effectiveness. In this way it may also be possible to reduce the capsule size, which will further improve patient compliance.

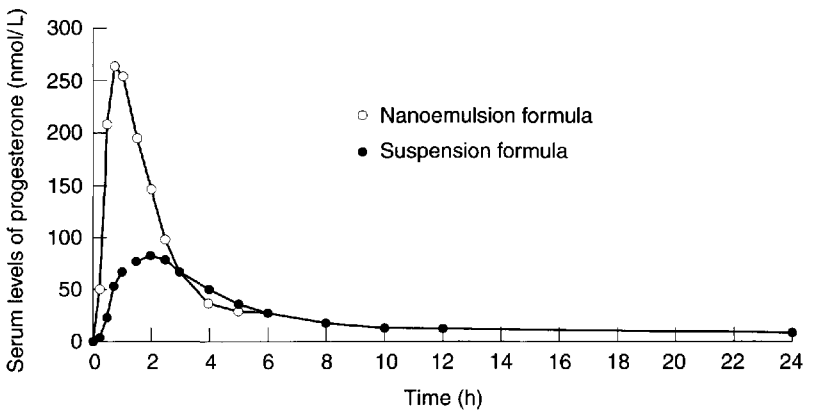


Fig. 30.5 Pharmacokinetic evaluation of progesterone comparing a softgel nanoemulsion solution of progesterone with a softgel containing a suspension of the drug in an oil following single-dose administration in 12 healthy human volunteers (Ferdinando 2000).

Safety for potent and cytotoxic drugs

The mixing, granulation and compression/filling processes used in preparing tablets and hard-shell capsules can generate a significant quantity of airborne powders. This can be of great concern for highly potent or cytotoxic compounds in terms of the operator and environmental protection required for satisfactorily safe product manufacture.

By preparing a solution or suspension of drug, where the active component is essentially protected from the environment by the liquid, such safety concerns can be significantly reduced.

Oils and low melting-point drugs

When the pharmaceutical active is an oily liquid, has a melting point less than about 75°C or proves difficult to compress, liquid filling of softgels is an obvious approach to presenting a solid oral dosage form. If the drug is an oily liquid, then it can be encapsulated directly into a softgel without adding a further diluent. Other low melting-point drugs may be formulated with a diluent oil in order to ensure satisfactory liquid flow and dosing into softgels.

Dose uniformity of low-dose drugs

In pharmaceutical manufacture liquid dosing avoids the difficulties of poor powder flow and therefore poor content uniformity. This is an important benefit for formulations containing drug doses in the microgram region. Attempts to produce adequate mixtures of small quantities of a low-dose drug in larger quantities of powdered excipients for tabletting or hard-shell filling are often unsatisfactory. In contrast, improved homogeneity is achieved by dissolving the drug in a liquid and then encapsulating the liquid matrix in a softgel.

Product stability

If a drug is subject to oxidative or hydrolytic degradation, the preparation of a liquid-filled softgel may prove beneficial. The liquid is prepared and encapsulated under a protective nitrogen atmosphere and the subsequently dried shell has very low oxygen permeability. By formulating in a lipophilic vehicle and packaging in well designed blister packs using materials of low moisture transmission, the drug can be protected from moisture. Conversely, it is well accepted that, in a solution, the drug may be more reactive than in the dry state and therefore potentially less stable. The appropriate choice of excipients, an understanding of the drug degradation pathways and appropriate preformulation studies are vital to achieving a stable product.

MANUFACTURE OF SOFTGELS

Softgel capsules were used in the 19th century as a means of administering bitter-tasting or liquid medicines. These were manufactured individually by preparing a small sack of gelatin and allowing it to set. Each sack, or gelatin shell, was then filled with the medication and heat-sealed. This method of manufacture was improved using a process that involved sealing two sheets of gelatin film between a pair of matching flat brass dies. Each die contained pockets into which the gelatin sheet was pressed and into which the medication was filled. The pressure between the two plates enabled individual capsules to be cut out from the die mould, and these capsules were subsequently dried.

However, it was not until the invention of the rotary die encapsulation machine by Robert Pauli Scherer in 1933 that liquid-fill capsules could be manufactured on a production scale. The rotary die process involves the continuous formation of a heat seal between two ribbons of gelatin, simultaneous with dosing of the fill liquid into each capsule. Although the speed and efficiency of the manufacturing process have improved greatly in recent years, the basic manufacturing principle remains essentially unchanged. The overall layout of a soft gelatin encapsulation machine is shown in Figure 30.6.

Before the encapsulation process takes place, there are two subprocesses that are often carried out simultaneously, yielding the two components of a softgel. These are (a) the gel mass which will provide the softgel shell, and (b) the fill matrix for the contents.

The gel mass is prepared by dissolving the gelatin in water at approximately 80°C and under vacuum, followed by the addition of the plasticizer, for example glycerol. Once the gelatin is fully dissolved then other components, such as colours, opacifier, flavours and preservatives, may be added. The hot gel mass is then supplied to the encapsulation machine through heated transfer pipes by a casting method that forms two separate gelatin ribbons, each approximately 150 mm wide. During the casting process the gelatin passes through the sol–gel transition and the thickness of each gel ribbon is controlled to ± 0.1 mm, in the range of about 0.5–1.5 mm. The thickness is checked regularly during the manufacturing process.

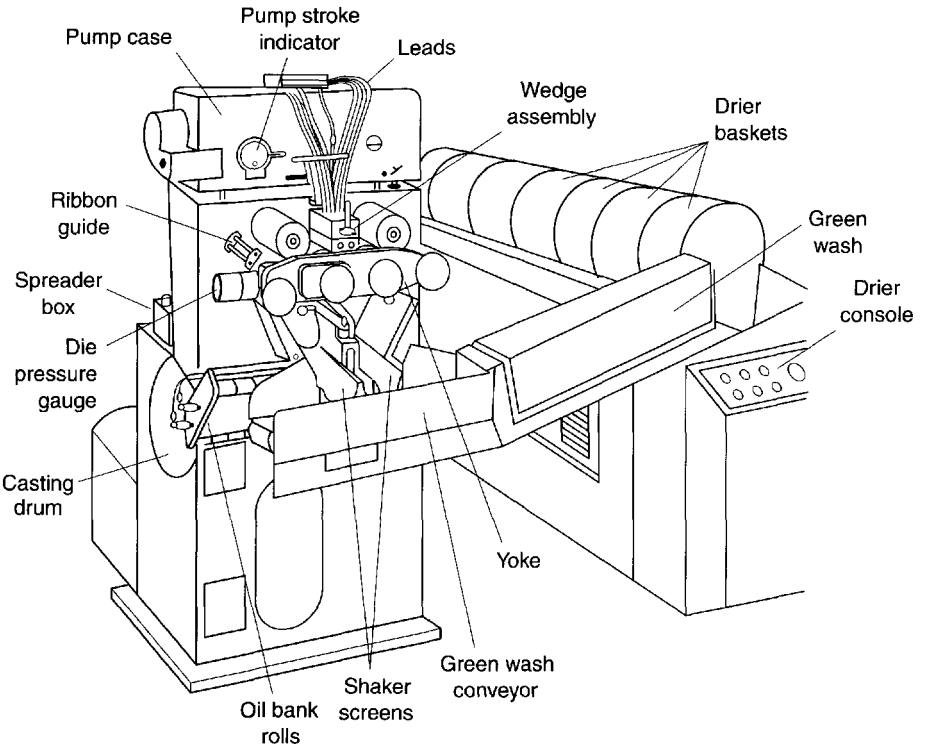


Fig. 30.6 Diagram of a soft gelatin encapsulation machine.

The two gel ribbons are then carried through rollers (at which a small quantity of vegetable oil lubricant is applied) and onwards to the rotary die encapsulation tooling (Fig. 30.7). Each ribbon provides one half of the softgel. It is possible to make bicoloured softgels using gel ribbons of two different colours.

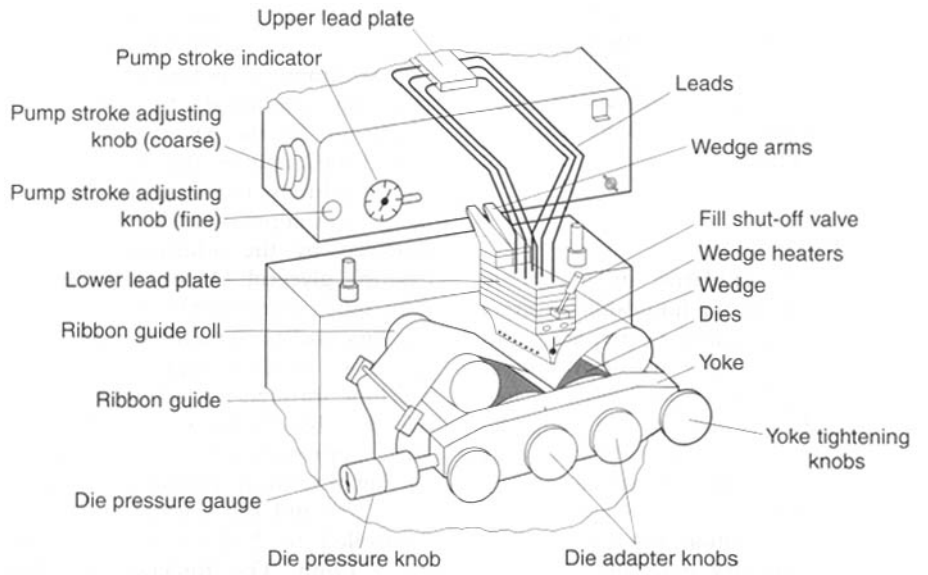


Fig. 30.7 Detail of a soft gelatin encapsulation machine.

The liquid fill matrix containing the active drug substance is manufactured separately from preparation of the molten gel. Manufacture of the active fill matrix involves dispersing or dissolving the drug substance in the non-aqueous liquid vehicle using conventional mixer–homogenizers.

A number of different parameters are controlled during the preparation of the active fill matrix, depending on the properties of the drug substance. For example, oxygen-sensitive drugs are protected by mixing under vacuum and/or inert gas; and in some cases an antioxidant component may be added to the formulation. Also, if the drug substance is present as a suspension in the liquid fill matrix then it is important to ensure that particle size of the drug does not exceed approximately 200 μm. By doing this it is possible to ensure that drug particles do not become trapped within the capsule seal, potentially leading to loss of integrity of the softgel.

Rotary die encapsulation is the process by which the gel ribbon and the unit dose of liquid fill matrix are combined to form the softgel. The process involves careful control of three parameters:

1. Temperature. This controls the heat available for capsule seal formation.
2. Timing. The timing of the dosing of unit quantities of fill matrix into the softgel during its formation is critical.
3. Pressure. The pressure exerted between the two rotary dies controls the softgel shape and the final cut-out from the gel ribbon.

Figure 30.8 is a simplified diagram representing the mechanism of softgel formation using contrarotating dies and the wedge-shaped fill-matrix injection system.

Accurately metered volumes of the liquid fill matrix are injected from the wedge device into the space between the gelatin ribbons as they pass between the die rolls. The wedge-shaped injection system is itself heated to approximately 40°C. The injection of liquid between the ribbons forces the gel to expand into the pockets of the dies, which govern the size and shape of the softgels. The ribbon continues to flow past the heated wedge injection system and is then pressed between the die rolls. Here the two softgel capsule halves are sealed together by the application of heat and pressure. The capsules are cut automatically from the gel ribbon by raised rims around each die on the rollers.

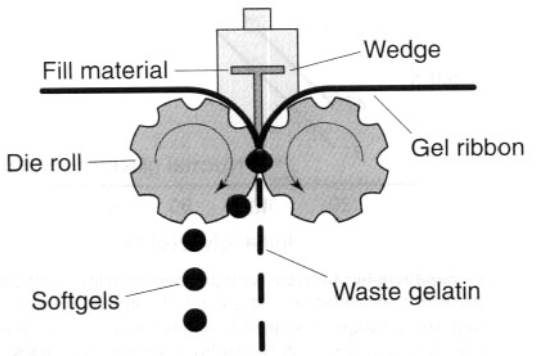


Fig. 30.8 Softgel formation mechanism.

After manufacture the capsules are passed through a tumble drier and then, to complete the drying process, spread on to trays and stacked in a tunnel drier that supplies air at 20% relative humidity. The tunnel drying process may take 2 or 3 days, or possibly as long as 2 weeks, depending on the specific softgel formulation. Finally, the softgels are inspected and packed into bulk containers in order to prevent further drying and for storage.

FORMULATION OF SOFTGELS

Gelatin shell formulation

Typical softgel shells are made up of gelatin, plasticizer, and materials that impart the desired appearance (colourants and/or opacifiers), and sometimes flavours. The following sections describe each of these materials, their functions, types, and amounts most often used in manufacturing softgel shells.

Gelatin

A large number of different gelatin shell formulations are available, depending on the nature of the liquid fill matrix. Most commonly the gelatin is alkali- (or base-) processed (type B) and it normally constitutes 40% of the wet molten gel mass. Type A acid-processed gelatin can also be used.

Plasticizers

Plasticizers are used to make the softgel shell elastic and pliable. They usually account for 20–30% of the wet gel formulation. The most common plasticizer used in softgels is glycerol, although sorbitol and propylene glycol are also frequently used, often in combination with glycerol. The amount and choice of the plasticizer contribute to the hardness of the final product and may even affect its dissolution or disintegration characteristics, as well as its physical and chemical stability. Plasticizers are selected on

the basis of their compatibility with the fill formulation, ease of processing, and the desired properties of the final softgel, including hardness, appearance, handling characteristics and physical stability.

One of the most important aspect of softgel formulation is to ensure that there is minimum interaction or migration between the liquid fill matrix and the softgel shell. The choice of plasticizer type and concentration is important in ensuring optimum compatibility of the shell with the liquid fill matrix.

Water

The other essential component of the softgel shell is water. Water usually accounts for 30–40% of the wet gel formulation and its presence is important to ensure proper processing during gel preparation and softgel encapsulation. Following encapsulation, excess water is removed from the softgels through controlled drying. In dry softgels the equilibrium water content is typically in the range 5–8% w/w, which represents the proportion of water that is bound to the gelatin in the softgel shell. This level of water is important for good physical stability, because in harsh storage conditions softgels will become either too soft and fuse together, or too hard and embrittled.

Colourants/opacifiers

Colourants (soluble dyes, or insoluble pigments or lakes) and opacifiers are typically used at low concentrations in the wet gel formulation. Colourants can be either synthetic or natural, and are used to impart the desired shell colour for product identification. An opacifier, usually titanium dioxide, may be added to produce an opaque shell when the fill formulation is a suspension, or to prevent photodegradation of light-sensitive fill ingredients. Titanium dioxide can either be used alone to produce a white opaque shell or in combination with pigments to produce a coloured opaque shell.

Properties of soft gelatin shells

Oxygen permeability

The gelatin shell of a softgel capsule provides a good barrier against the diffusion of oxygen into its contents. The quantity of oxygen *(q)* that passes through the gelatin is governed by the area *(A)*, thickness *(h)*, particle pressure difference $(p_1 - p_2)$, time of diffusion *(t)* and the permeability coefficient *(P)* of the shell by the following equation:

$$q = \frac{PAt\,(p_1 - p_2)}{h} \qquad (30.1)$$

The permeability coefficient *(P)* is related to the diffusion coefficient *(D)* and the solubility coefficient *(S)* by the equation $P = DS$. This relationship, described by Henry's Law, assumes no interaction between the gas and the polymeric film, but *P* is clearly affected by the formulation of the gelatin shell, as shown in Figure 30.9.

Figure 30.9 shows the relationship between oxygen permeability coefficient and the glycerol concentration in the gelatin shell of softgels at room temperature and relative humidity values from 31 to 80%. The oxygen permeability decreases with the % RH and the glycerol content in the gelatin shell formulation. For maximum protection against the ingress of oxygen, the gelatin shell should be dry and formulated to contain about 30–40% glycerol.

Residual water content

Softgels contain little residual water, and compounds which are susceptible to hydrolysis are protected if dissolved or dispersed in an oily liquid fill material and encapsulated as a soft gelatin capsule. Figure 30.10 shows the relationship between the equilibrium water content and the concentration of glycerol in the gelatin shell of a softgel capsule, stored at room temperature and environmental relative humidities of between 31 and 80%. The data show, for example, that the minimum water values are found at glycerol levels in the shell of

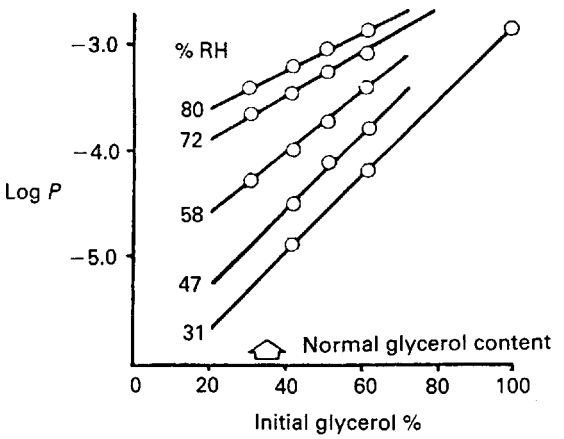


Fig. 30.9 Relationship between oxygen permeability coefficient and the glycerol concentration in the shell of softgels at room temperature and a range of relative humidity values. From Hom, F.S., Veresh, S.A. and Ebert, W.R. (1975) *J. Pharm. Sci.*, **64(5)**, 851–857.

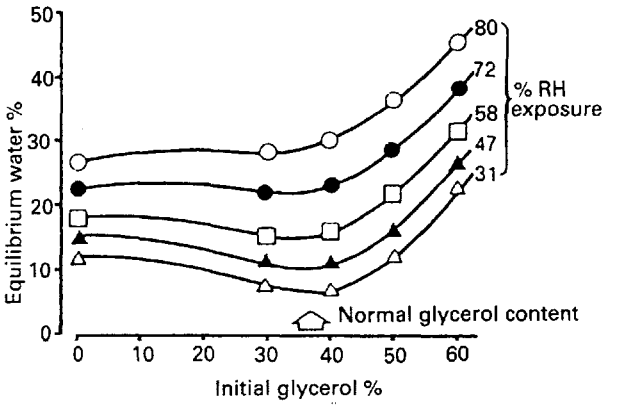


Fig. 30.10 The relationship between equilibrium water content and the concentration of glycerol in the shell of soft gelatin capsules at room temperature and a range of relative humidity values. (See Hom et al (1975), as Fig. 30.9.)

between 30 and 40%. Such a formulation dried at 31% relative humidity has a water content in the shell of about 7% and a water content in the fill in equilibrium with the atmosphere. The residual water content of most pharmaceutical compounds stored at 20% relative humidity (the drying condition for softgels) is low, and the water levels in the fills of softgels are therefore very small.

Formulation of softgel fill materials

In terms of formulation requirements, the softgel should be considered as a biphasic dosage form: a solid-phase capsule shell and a liquid-phase fill matrix. Although it is possible to incorporate a drug in the shell of a softgel, the overwhelming majority of products have the active ingredient(s) within the fill matrix. The liquid-phase fill matrix is selected from components with a wide range of different physicochemical properties. The choice of components is made according to one or more of a number of criteria, including the following:

- Capacity to dissolve the drug;
- Rate of dispersion in the gastrointestinal tract after the softgel shell ruptures and releases the fill matrix;
- Capacity to retain the drug in solution in the gastrointestinal fluid;
- Compatibility with the softgel shell;
- Ability to optimize the rate, extent and consistency of drug absorbed.

Types of softgel fill matrices

Lipophilic liquids/oils Triglyceride oils, such as soya bean oil, are commonly used in softgels. When used alone, however, their capacity to dissolve drugs is limited. Nevertheless, active ingredients such as hydroxycholecalciferol and other vitamin D analogues, plus steroids such as oestradiol, can be formulated into simple oily solutions for encapsulation in softgels.

Hydrophilic liquids Polar liquids with a sufficiently high molecular weight are commonly used. Polyethylene glycol (PEG) is the most frequently used, for example PEG 400, which has an average molecular weight of approximately 400 Da. Smaller hydrophilic molecules, such as ethanol or indeed water, can be incorporated in the softgel fill matrix in low levels, typically below 10% by weight.

Self-emulsifying oils A combination of a pharmaceutical oil and a non-ionic surfactant such as polyoxyethylene sorbitan mono-oleate can provide an oily formulation which disperses rapidly in the gastrointestinal fluid. The resulting oil/surfactant droplets enable rapid transfer of the drug to the absorbing mucosa and subsequent drug absorption.

Microemulsion and nanoemulsion systems A microemulsion of a lipid-surfactant–polar liquid system is characterized by its translucent single-phase appearance. The droplet size is in the submicrometre range, and light scattering by these droplets results in a faint blue colouration known as the Tyndall effect.

A nanoemulsion is a similar system but contains emulsion droplets in the 100 nm size range. Microemulsion and nanoemulsion systems have the advantage of a high capacity to solubilize drug compounds and to retain the drug in solution even after dilution in gastrointestinal fluids.

In order to produce a microemulsion or nanoemulsion in the gastrointestinal tract a 'preconcentrate' is formulated in the softgel fill matrix. In other words, the preconcentrate fill matrix contains a lipid component and one or more surfactants, which

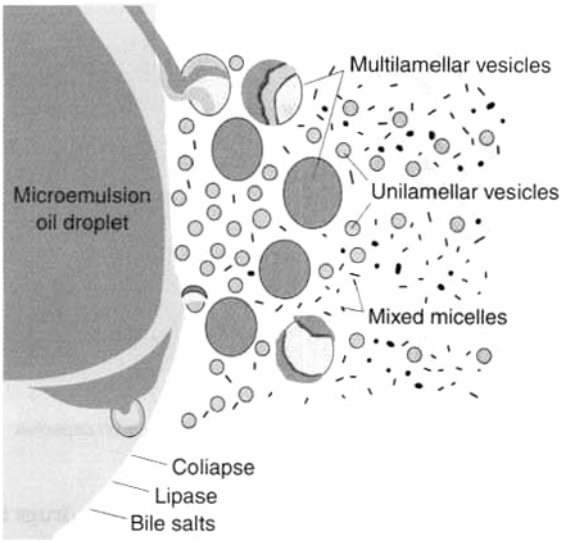


Fig. 30.11 Diagram of proposed nanoemulsion/microemulsion dissolution mechanism.

spontaneously form a microemulsion or a nanoemulsion on dilution in an aqueous environment, such as in gastrointestinal fluid (Fig. 30.11). The resulting microemulsion or nanoemulsion is often stable for prolonged periods.

Suspensions Drugs that are insoluble in softgel fill matrices are formulated as suspensions. The continuous phase may be any of the vehicles described above. Suspension formulations provide significant advantages for certain low-solubility drugs which are very poorly absorbed after oral administration. With the appropriate choice of excipients, softgel suspensions can have improved bioavailability compared to compressed tablets or hard-shell capsules, or even dilute aqueous solutions.

Lipolysis systems The advantage of the microemulsion approach lies in the high surface area presented by the microemulsion particles, which are essentially surfactant micelles swollen with solubilized oil and drug. This high surface area facilitates the rapid diffusion of drug from the dispersed oil phase into the aqueous intestinal fluids, until an equilibrium distribution is established. Thereafter, as drug is removed from the intestinal fluids via enterocyte absorption, it is quickly replenished by the flow of fresh material from the microemulsion particles.

Formulation using the lipolysis systems

In addition to promoting the solubility of drug compounds, lipid formulations can also facilitate dissolution by taking advantage of lipolysis. This is because the lipid components of a softgel fill matrix, which comprise triglycerides or a partial (mono-/di-) glyceride, are often subject to intestinal fat digestion or lipolysis. Lipolysis is the term used to describe the action of the enzyme pancreatic lipase on triglycerides and partial glycerides, to form 2-monoglycerides and fatty acids. These 2-monoglycerides and fatty acids, known as lipolytic products, then interact with bile salts to form small droplets, or vesicles. These vesicles are broken down into smaller and smaller vesicles, ultimately resulting in the formation of mixed micelles that are approximately 3–10 nm in size.

If a drug compound possesses higher solubility in lipolytic products than in triglyceride oils, then it is advantageous for lipolysis to occur in the intestinal lumen. In this way, the process of lipolysis promotes the formation of an excellent dissolution medium for the drug, namely lipolytic products. On the other hand, the absorption of a drug compound may be adversely affected by the presence of bile salt, and in such a case it may be advantageous for lipolysis to be reduced or blocked completely. It has been found that certain hydrophilic and lipophilic surfactants have the ability to block or promote lipolysis respectively (MacGregor et al 1997). These hydrophilic and lipophilic surfactants are often used in softgel fill matrix formulations.

It is possible to measure the rate and extent of lipolysis for a softgel fill matrix formulation. This is done by an in vitro pH stat measurement technique. The experimental conditions for this model are as shown in Table 30.2.

In this model, lipolysis is quantified by the amount of free fatty acids liberated by enzymatic digestion of the lipids in the softgel fill matrix. The quantity of 1.0 M sodium hydroxide titrant is directly proportional to the extent of lipolysis.

The mixed intestinal micelles produced as a result of this lipolysis process are physiologically important

Table 30.2 An in vitro lipolysis model

- 100 mL pH 6.5 *tris*-maleate buffer at 37°C
- 2.4 % w/w triglyceride oil in a softgel fill matrix formulation
- 20 TBUs* per mL of porcine pancreatic lipase–colipase

  * 1 TBU (tributyrin unit) = 1 μmol butyric acid per minute
- 8 mM bile salt (sodium taurocholate)
- 1.5 mM lecithin, 5 mM $CaCl_2 \cdot 2H_2O$, 150 mM NaCl
- pH maintained at 6.5 using 1.0 M NaOH via continuous titration (pH-stat)

because they can transport high concentrations of hydrophobic molecules across the aqueous boundary layer that separates the absorptive membrane from the intestinal lumen. Thus, lipolytic products (e.g. fatty acids and monoglycerides) – and hydrophobic drug, if present – reside in the hydrophobic core regions of mixed intestinal micelles. In contrast, the surface of the micelles remains hydrophilic and this facilitates rapid micellar diffusion across the aqueous boundary layer to the intestinal membrane. In the microclimate adjacent to the intestinal membrane the pH is lower than in the intestinal lumen. This promotes demicellization, leading to the formation of a supersaturated solution of lipolytic products (and hydrophobic drug, if present) in close proximity to the enterocyte surface. These materials are then readily absorbed across the cell membrane by passive diffusion.

Mixed intestinal micelles comprising bile salts and lipolytic products can enhance the bioavailability of hydrophobic drugs whose absorption is normally dissolution-rate limited. This is because mixed intestinal micelles can be very potent solubilizing agents for a wide range of hydrophobic drugs, much more so than simple bile salt micelles formed in the absence of lipolytic products. For example, under simulated physiological conditions the aqueous solubility of cinnarizine in simple bile salt micelles is 4 μg/mL, compared to 0.5 μg/mL in aqueous buffer. However, in the presence of mixed micelles the solubility of cinnarizine is further enhanced to approximately 44 μg/mL (Embleton et al 1995).

Taking cinnarizine as an example, it would be advantageous to formulate a softgel fill matrix that allows lipolysis to occur in the intestinal lumen because of the high drug solubility in lipolytic products. If the inhibition by a hydrophobic surfactant were allowed to occur, then it is highly likely that cinnarizine absorption would be impaired because of the reduced flow of drug into mixed micelles. However, if certain lipophilic surfactants, with a HLB less than 10, are added to the formulation, then the inhibitory effects of hydrophilic surfactants on lipolysis can be reduced or eliminated.

Two formulations containing cinnarizine, a hydrophobic drug whose absorption is normally dissolution-rate limited, have been compared (Embleton et al 1995). Formulation [A] was prepared as a lipolysing formulation and [B] as a non-lipolysing formulation, as demonstrated by the in vitro model. Formulation [A] was composed of a digestible triglyceride oil, a hydrophilic surfactant and a lipophilic surfactant, which was chosen for its ability to overcome the inhibitory effects of the hydrophilic surfactant on the in vitro triglyceride lipolysis. In vitro this formulation exhibited 79% lipolysis after 60 minutes, compared to the digestible oil alone. In contrast, the non-lipolysing formulation contained a lipophilic surfactant that did not overcome the inhibitory effects of the hydrophilic surfactant on the lipolysis of the triglyceride oil, and was shown to lipolyse to an extent of only 3%. It is proposed that the oil in formulation [A], which forms a fine oil-in-water emulsion on aqueous dilution, is rapidly digested, forming mixed intestinal micelles with endogenous bile. These micelles transport the drug to the intestinal membrane, where the pH of the microclimate promotes micellar breakdown, facilitating enterocyte transport to the systemic circulation. In contrast, on dilution with aqueous fluids, formulation [B] forms a translucent microemulsion (as indicated by a blue tinge resulting from the Tyndall effect). As a result of this formulation failing to lipolyse and thereby remaining unaffected by enzymic activity, the drug is maintained within the oil phase, inhibiting the production of mixed intestinal micelles and restricting drug absorption.

The significance of the lipolysis process in enhancing the bioavailability of hydrophobic drugs was investigated further with an in vivo study (Fig. 30.12). This compared the bioavailability of cinnarizine (30 mg) administered orally as the lipolysing formulation [A] and non-lipolysing formulation [B] with a commercially available tablet, formulation [C], to six beagle dogs. The AUC(0–24 h) for formulation [A] was significantly increased by 64% ($P < 0.01$) compared to the tablet preparation, and by 48% ($P < 0.001$) compared to formulation [B]. The C_{max} of formulation [A] was approximately 75% higher than both formulations [B], ($P < 0.001$) and [C] ($P < 0.01$).

The results of this study have given a valuable insight into the effect of microemulsion formulation on the absorption of a hydrophobic drug in the gastrointestinal tract, and new information as to how the lipolysis process may influence bioavailability (Lacy et al 2000).

PRODUCT QUALITY CONSIDERATIONS

Ingredient specifications

All of the ingredients of a softgel are controlled and tested to ensure compliance with pharmacopoeial specifications. However, additional specification tests may be added for certain excipients in order to ensure the manufacture of a high-quality softgel product. For example, it is important to limit certain trace impurities, such as aldehydes and peroxides

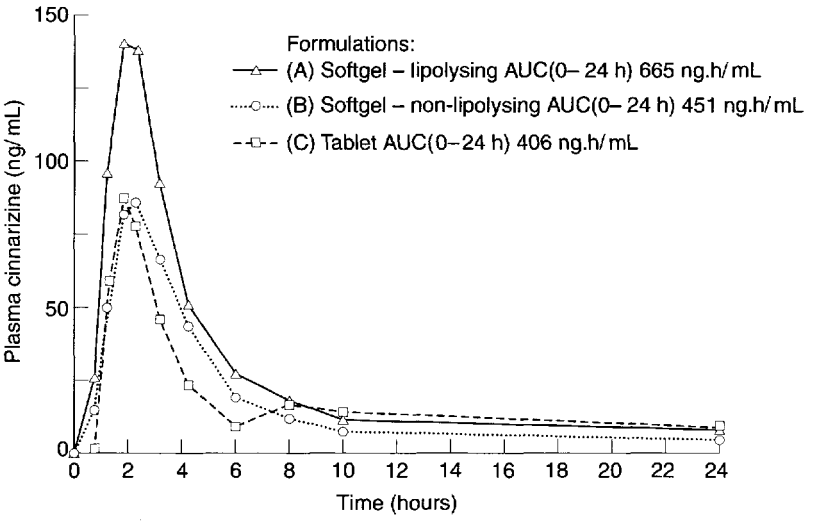


Fig. 30.12 Plasma concentration versus time curves for three formulations of cinnarizine in the dog (n=6) (Embleton 1995).

which may be present in polyethylene glycol. The presence of high levels of these impurities gives rise to cross-linking of the gelatin polymer, leading to insolubilization through further polymerization. On prolonged storage this can lead to slow dissolution of the capsule shell and subsequent retarded drug release.

The ingredient requiring the most careful control is gelatin itself. Once a particular grade of gelatin is used in a softgel formulation the quality is controlled using parameters such as the viscosity of a hot solution and the bloom strength of the gel (bloom strength is a measure of the hardness of a gel).

In-process testing

During the encapsulation process the four most important tests are:

- The gel ribbon thickness;
- Softgel seal thickness at the time of encapsulation;
- Fill matrix weight and capsule shell weight;
- Softgel shell moisture level and softgel hardness at the end of the drying stage.

Appropriate control levels for these parameters are established during process development for each softgel product, and are applied in routine production scale manufacture.

Finished product testing

Finished softgels are subjected to a number of tests in accordance with compendial requirements for unit dose capsule products. These normally include capsule appearance, active ingredient assay and related substances assay, as well as fill weight, content uniformity and microbiological testing.

REFERENCES

Aungst B.J. (2000) Mini Review Intestinal Permeation Enchancers. *J. Pharm. Sci.* **89(4)**: 429–442.

Drewe J., Meier R., Vonderscherer J., et al. (1992) Enhancement of the oral absorption of cyclosporin in man. *Br. J. Clin. Pharmacol.* **34**: 60–64.

Embleton J., Hutchison K.G., Lacy J., et al. (1995) The effect of in-vivo lipolysis in improving the bioavailability of orally administered cinnarizine in self-emulsifying oily vehicles. American Association of Pharmaceutical Sciences Conference, Miami Beach.

Ferdinando J.C. (2000) Formulation Solutions – Softgels. Manufacturing and Packaging; 69–73.

Lacy J.E., Embleton J.K., Perry E.A. (2000) Delivery systems for hydrophobic drugs: US Patent 6,096,338.

MacGregor K.J., Embleton J.K., Lacy J.E, et al. (1997) Influence of lipolysis on drug absorption from the gastrointestinal tract. *Advd Drug Del. Rev.* **25**: 33–46.

Meinzer A. (1993) Sandimmun Neoral Soft Gelatin Capsules. International Industrial Pharmaceutical Research Conference. Wisconsin USA.

Perry C.M., Noble S. (1998) Saquinavir Softgel Capsule Formulation. *Drugs* **55:(3)**.

Saano V., Paronen P., Peura P., et al. (1991) Relative pharmacokinetics of three oral 400 mg ibuprofen dosage forms in healthy volunteers. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **29(10)**: 381–385.

31
Pulmonary drug delivery

Kevin Taylor

CHAPTER CONTENTS

INHALED DRUG DELIVERY

Drugs are generally delivered to the respiratory tract for the treatment or prophylaxis of airways diseases, such as bronchial asthma and cystic fibrosis. The administration of a drug at its site of action can result in a rapid onset of activity, which may be highly desirable, for instance when delivering bronchodilating drugs for the treatment of asthma. Additionally, smaller doses can be administered locally compared to delivery by the oral or parenteral routes, thereby reducing the potential incidence of adverse systemic effects and reducing drug costs. The pulmonary route is also useful where a drug is poorly absorbed orally, e.g. sodium cromoglycate, or where it is rapidly metabolized orally, e.g. isoprenaline. The avoidance of first-pass metabolism in the liver may also be advantageous, although the lung itself has some metabolic capability.

The lung may be used as a route for delivering drugs having systemic activity, because of its large surface area, the abundance of capillaries and the thinness of the air–blood barrier. This has been exploited in the treatment of migraine with ergotamine, and studies have demonstrated the potential for delivering proteins and peptides such as insulin and growth hormone via the airways.

Lung anatomy

The lung is the organ of external respiration, in which oxygen and carbon dioxide are exchanged between blood and inhaled air. The structure of the airways also prevents the entry of and promotes efficient removal of airborne foreign particles, including microorganisms.

The respiratory tract can be considered as comprising conducting (central) regions (trachea, bronchi, bronchioles, terminal and respiratory bronchioles) and respiratory (peripheral) regions

(respiratory bronchioles and alveolar regions), although there is no clear demarcation between them (Fig. 31.1). The upper respiratory tract comprises the nose, throat, pharynx and larynx; the lower tract comprises the trachea, bronchi, bronchioles and the alveolar regions. Simplistically, the airways can be described by a symmetrical model in which each airway divides into two equivalent branches or generations. In fact, the trachea (generation 0) branches into two main bronchi (generation 1), of which the right bronchus is wider and leaves the trachea at a smaller angle than the left, and hence is more likely to receive inhaled material. Further branching of the airways ultimately results in terminal bronchioles. These divide to produce respiratory bronchioles, which connect with alveolar ducts leading to the alveolar sacs (generation 23). These contain approximately $2–6 \times 10^8$ alveoli, producing a surface area of about 70–80 m^2 in an adult male.

The conducting airways are lined with ciliated epithelial cells. Insoluble particles deposited on the airways walls in this region are trapped by the mucus and swept upwards from the lungs by the beating cilia and swallowed.

Inhalation aerosols and the importance of size distribution

To deliver a drug into the airways it must be presented as an **aerosol**. In pharmacy this is defined as a two-phase system of solid particles or liquid droplets dispersed in air or other gaseous phase, having sufficiently small size to display considerable stability as a suspension.

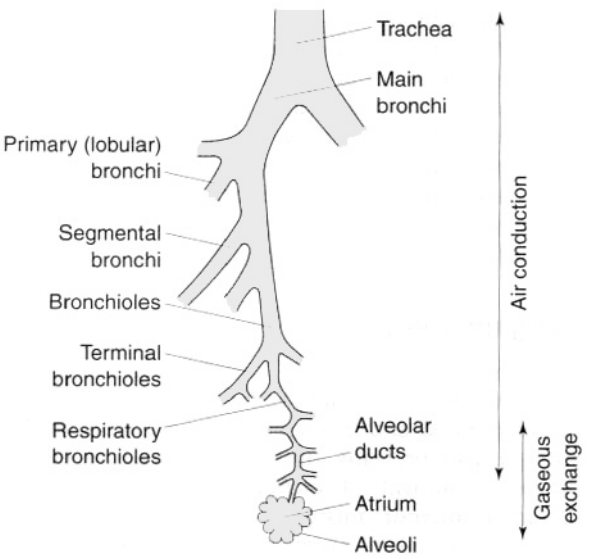


Fig. 31.1 Schematic representation of the human airways. (Reproduced with permission from Wilson and Washington 1989.)

The deposition of a drug/aerosol in the airways is dependent on four factors: the physicochemical properties of the drug, the formulation, the delivery/liberating device and the patient (breathing patterns and clinical status).

The most fundamentally important physical property of an aerosol for inhalation is its size. The particle size of an aerosol is usually standardized by calculation of its ***aerodynamic diameter***, d_a, which is the physical diameter of a unit density sphere which settles through air with a velocity equal to the particle in question. Therapeutic aerosols are heterodispersed (polydispersed), and the distribution of sizes is generally represented by the geometric standard deviation (GSD or σ_g), when size is log-normally distributed.

For approximately spherical particles

$$d_a = d_p \times \sqrt{\rho/\rho_o} \qquad (31.1)$$

where d_p is physical diameter, ρ is particle density and ρ_o is unit density, i.e. 1 g/cm^3.

When d_p is the mass median diameter (MMD), d_a is termed the mass median aerodynamic diameter (MMAD).

The influence of environmental humidity on particle size

As a particle enters the respiratory tract, the change from ambient to high relative humidity (approximately 99%) results in condensation of water on to the particle surface, which continues until the vapour pressure of the water equals that of the surrounding atmosphere. For water-insoluble materials this results in a negligibly thin film of water; however, with water-soluble materials a solution is formed on the particle surface. As the vapour pressure of the solution is lower than that of pure solvent at the same temperature, water will continue to condense until an equilibrium between vapour pressures is reached, i.e. the particle will increase in size. The final equilibrium diameter reached is constrained by the Kelvin effect, i.e., the vapour pressure of a droplet solution is higher than that for a planar surface, and is a function of the particle's original diameter (Pritchard, 1987). Hygroscopic growth will affect the deposition of particles, resulting in deposition higher in the respiratory tract than would have been predicted from measurements of their initial size.

Particle deposition in the airways

The efficacy of a clinical aerosol is dependent on its ability to penetrate the respiratory tract. To penetrate to the peripheral (respirable) regions, aerosols require a size less than about 5 or 6 μm, with less than 2 μm being preferable for alveolar deposition. Literature values for 'respirable' size vary and must be considered alongside the environmental changes in size described above and the heterodispersed nature of inhalation aerosol size distributions. Larger particles or droplets are deposited in the upper respiratory tract and are rapidly cleared from the lung by the mucociliary action, with the effect that drug becomes available for systemic absorption and may potentially cause adverse effects. Steroid aerosols of sufficiently large size may deposit in the mouth and throat, with the potential to cause oral candidiasis. The size of aerosolized drug may be especially important in the treatment of certain conditions where penetration to the peripheral airways is particularly desirable, for instance the treatment and prophylaxis of the alveolar infection *Pneumocystis carinii* pneumonia.

There are three main mechanisms responsible for particulate deposition in the lung: gravitational sedimentation, impaction and diffusion.

Gravitational sedimentation

From Stokes' law, particles settling under gravity will attain a constant terminal settling velocity, U_t:

$$U_t = \frac{\rho g d^2}{18\eta} \qquad (31.2)$$

where ρ is particle density, g is the gravitational constant, d is particle diameter and η is air viscosity.

Thus, gravitational sedimentation of an inhaled particle is dependent on its size and density, in addition to its residence time in the airways. Sedimentation is an important deposition mechanism for particles in the size range 0.5–3 μm, in the small airways and alveoli, for particles that have escaped deposition by impaction.

Inertial impaction

Where a bifurcation occurs in the respiratory tract, the airstream changes direction and particles within the airstream, having sufficiently high momentum, will impact on the airways' walls rather than follow the changing airstream. This deposition mechanism is particularly important for large particles having a diameter greater than 5 μm, and particularly greater than 10 μm, and is common in the upper airways, being the principal mechanism for deposition in the nose, mouth, pharynx and larynx and the large conducting airways. With the continuous branching of the conducting airways, the velocity of the airstream decreases and impaction becomes a less important mechanism for deposition.

The probability of impaction is proportional to:

$$\frac{U_t U \sin\theta}{gr} \qquad (31.3)$$

where θ is the change in airways direction, U is airstream velocity and r is the airway's radius.

Brownian diffusion

Collision and bombardment of small particles by molecules in the respiratory tract produce Brownian motion. The resultant movement of particles from high to low concentrations causes them to move from the aerosol cloud to the airways walls. The rate of diffusion is inversely proportional to the particle size, and thus diffusion is the predominant mechanism for particles smaller than 0.5 μm.

Other methods of deposition

Although impaction, sedimentation and diffusion are the most important mechanisms for drug deposition in the respiratory tract, other mechanisms may occur. These include ***interception***, whereby particles having extreme shapes, such as fibres, physically catch on to the airways' walls as they pass through the respiratory tract, and ***electrostatic attraction***, whereby an electrostatic charge on a particle induces an opposing charge on the walls of the respiratory tract, resulting in attraction between particle and walls.

Different deposition mechanisms are important for differently sized particles. Those greater than 5 μm will deposit predominantly by inertial impaction in the upper airways. Particles sized between 1 and 5 μm deposit predominantly by gravitational sedimentation in the lower airways, especially during slow, deep breathing, and particles less than 1 μm deposit by Brownian diffusion in the stagnant air of the lower airways. Particles of approximately 0.5 μm are inefficiently deposited, being too large for effective deposition by Brownian diffusion and too small for effective impaction or sedimentation, and they are often immediately exhaled. This ***size of minimum deposition*** should thus be considered during formulation, although for the reasons of environmental humidity discussed previously, the equilibrium diameter in the airways may be significantly larger than the original particle size in the formulation.

Breathing patterns

Patient-dependent factors, such as breathing patterns and lung physiology, also affect particle deposition. For instance, the larger the inhaled volume the greater the peripheral distribution of particles in the lung, whereas increasing inhalation flow rate enhances deposition in the larger airways by inertial impaction. Breath-holding after inhalation enhances the deposition of particles by sedimentation and diffusion. Optimal aerosol deposition occurs with slow, deep inhalations to total lung capacity, followed by breath-holding prior to exhalation. It should be noted that changes in the airways resulting from disease states, for instance airways' obstruction, may affect the deposition profile of an inhaled aerosol.

Clearance of inhaled particles and drug absorption

Particles deposited in the ciliated conducting airways are cleared within 24 hours and ultimately swallowed. Insoluble particles penetrating to the alveolar regions, and which are not solubilized in situ, are removed more slowly. Alveolar macrophages engulf such particles and may then migrate to the bottom of the mucociliary escalator, or alternatively particles may be removed via the lymphatics.

Hydrophobic compounds are usually absorbed at a rate dependent on their oil/water partition coefficients, whereas hydrophilic materials are poorly absorbed through membrane pores at rates inversely proportional to molecular size. Thus, the airways' membrane, like the gastrointestinal tract, is preferably permeable to the unionized form of a drug. Some drugs, such as sodium cromoglycate, are partly absorbed by a saturable active transport mechanism. The rate of drug absorption, and consequently drug action, can be influenced by the formulation. Rapid drug action can generally be achieved using solutions or powders of aqueous soluble salts, whereas slower or prolonged absorption may be achieved using suspension formulations, powders of less soluble salts, or novel drug delivery systems such as liposomes and microspheres.

FORMULATING AND DELIVERING THERAPEUTIC INHALATION AEROSOLS

There are currently three main types of aerosol-generating device for use in inhaled drug therapy: metered-dose inhalers, dry powder inhalers and nebulizers.

Metered-dose inhalers

Metered-dose inhalers (MDIs), introduced in the mid-1950s, are the most commonly used inhalation drug delivery devices. In MDIs, drug is either dissolved or suspended in a liquid propellant mixture together with other excipients, including surfactants, and presented in a pressurized canister fitted with a metering valve (Fig. 31.2). A predetermined dose is released as a spray on actuation of the metering valve. When released from the canister the formulation undergoes volume expansion in the passage within the valve and forms a mixture of gas and liquid before discharge from the orifice. The high-speed gas flow helps to break up the liquid into a fine spray of droplets.

Containers

Pharmaceutical aerosols may be packaged in tin-plated steel, plastic-coated glass or aluminium containers. In practice, MDIs are generally presented in aluminium canisters, produced by extrusion to give seamless containers with a capacity of 10–30 mL. Aluminium is relatively inert and may be used uncoated where there is no chemical instability between container and contents. Alternatively, aluminium containers with an internal coating of a chemically resistant organic material, such as an epoxy resin, can be used.

Propellants

The propellants used in MDI formulations are liquefied gases, traditionally chlorofluorocarbons (CFCs) and increasingly hydrofluoroalkanes (HFAs). At room temperature and pressure these are gases, but they are readily liquefied by decreasing

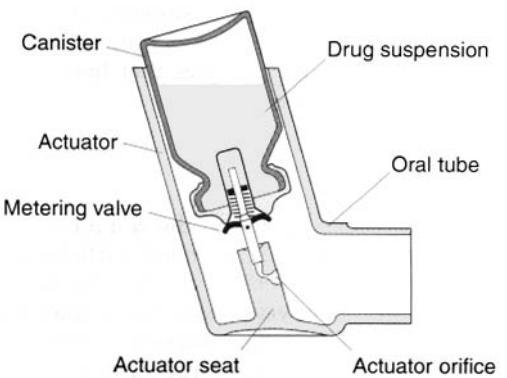


Fig. 31.2 The metered-dose inhaler. (Reproduced with permission from Morén 1981.)

temperature or increasing pressure. The head space of the aerosol is filled with propellant vapour, producing the saturation vapour pressure at that temperature. On spraying, medicament and propellant are expelled and the head volume increases. To re-establish the equilibrium, more propellant evaporates and so a constant pressure system with consistent spray characteristics is produced. The CFCs currently employed in MDI formulations are trichlorofluoromethane (CFC-11), dichlorodifluoromethane (CFC-12) and dichlorotetrafluoroethane (CFC-114). Formulations generally comprise blends of CFC-11 and CFC-12 or CFC-11, CFC-12, and CFC-114 (Table 31.1.), together with a surfactant such as a sorbitan ester, oleic acid or lecithin, which acts as a suspending agent and lubricates the valve.

CFCs and HFAs are numbered using a universal system. The first digit is the number of carbon atoms minus 1 (omitted if zero), the second is the number of hydrogen atoms plus 1, and the third is the number of fluorine atoms. Chlorine fills any remaining valencies, given the total number of atoms required to saturate the compound. If asymmetry is possible, this is designated by a letter. The symmetrical isomer is assigned the number described above; of the asymmetrical isomers, that designated the letter a is the most symmetrical, b the next most symmetrical, and so on. The CFCs are perfectly miscible with each other and suitable blends give a useful intermediate vapour pressure, usually about 450 kPa. The vapour pressure of the mixture of propellants is given by Raoult's law, i.e. the vapour pressure of a mixed system is equal to the sum of the mole fraction of each component multiplied by its vapour pressure:

$$P = p_a + p_b \tag{31.4}$$

where P is the total vapour pressure of the system and p_a and p_b are the partial vapour pressures of the components, a and b:

$$p_a = x_a p_a^0 \tag{31.5}$$

$$p_b = x_b p_b^0 \tag{31.6}$$

where x_a and x_b are the mole fractions and p_a^0 and p_b^0 are the partial vapour pressures of components a and b, respectively.

The reaction of CFCs with the ozone in the earth's stratosphere, which absorbs ultraviolet radiation at 300 nm, and their contribution to global warming is a major environmental concern. CFCs pass to the stratosphere, where in the presence of UV they liberate chlorine, which reacts with ozone. The depletion of stratospheric ozone results in increased exposure to the UV-B part of the UV spectrum, resulting in a number of adverse effects, in particular an increased incidence of skin cancer. The Montreal Protocol of 1987 was a global ban on the production of the five worst ozone-depleting CFCs by the year 2000. This was amended in 1992, so that production of CFCs in developed countries was phased out by 1 January 1996. In the European Union, all ozone-depleting CFCs had been banned by the end of 1995. Pharmaceutical aerosols are currently exempted, but this exemption is reviewed annually. In household and cosmetic aerosols CFCs have been replaced by hydrocarbons such as propane and butane. Alternatively, compressed gases such as nitrogen dioxide, nitrogen and carbon dioxide may be used, for instance in food products. However, compressed gases do not maintain a constant pressure within the canister throughout its use, as the internal pressure is inversely proportionate to the head volume, and so product performance changes with age. For reasons of toxicity and inflammability, hydrocarbons are not considered appropriate alternatives to CFCs for inhalation products, and so non-ozone depleting alternatives to CFCs are being developed.

Propellants HFA-134a (trifluoromonofluoroethane) and HFA-227 (heptafluoropropane) are non-ozone depleting, non-flammable HFAs, also called hydrofluorocarbons (HFCs), which have been widely investigated as alternatives to CFC-12 (Table 31.2). However, these gases contribute to global warming and further replacements will no doubt be required in the future.

HFA-134a and HFA-227 have some physical properties, including density, which are similar to

Table 31.1 Formulae and physicochemical properties of chlorofluorocarbons (CFCs) used in MDI formulations

Number	Formula	Boiling point (°C)	Vapour pressure (kPa at 20°C)	Density (g/mL at 20°C)
11	CCl_3F	23.7	89 (0.89 bar)	1.49
12	CCl_2F_2	−29.8	568 (5.68 bar)	1.33
114	$C_2Cl_2F_4$	3.6	183 (1.83 bar)	1.47

Table 31.2 Formulae and physicochemical properties of hydrofluoroalkanes (HFAs) used in MDI formulations

Number	Formula	Boiling point (°C)	Vapour pressure (kPa at 20°C)	Density (g/mL at 20°C)
134a	$C_2F_4H_2$	−26.5	660 (6.6 bar)	1.23
227	C_3F_7H	−17.3	398 (3.98 bar)	1.41

those of CFC-12 and, to a lesser extent, CFC-114. However, they present major formulation problems: in particular they are poor solvents for the surfactants commonly used in MDI formulation and no alternative to CFC-11 is currently available. Ethanol is approved for use in formulations containing HFAs to allow dissolution of surfactants, and is included in marketed non-CFC MDI products. However, ethanol has low volatility and may consequently increase the droplet size of the emitted aerosols.

Metering valve

The metering valve of an MDI permits the reproducible delivery of small volumes (25–100 μL) of product. Unlike the non-metering continuous-spray valves of conventional pressurized aerosols, the metering valve in MDIs are used in the inverted position (Fig. 31.3). Depression of the valve stem allows the contents of the metering chamber to be discharged through the orifice in the valve stem and made available to the patient. After actuation, the metering chamber refills with liquid from the bulk and is ready to dispense the next dose. A corollary of this is that the MDI needs to be primed, i.e. the metering chamber filled, prior to the first use by a patient. MDI valves are complex in design and must protect the product from the environment, while also protecting against product loss during repeated use. The introduction of HFA propellants with different solvent properties has necessitated the development of new valve elastomers. The valve stem fits into the actuator, which is made of polyethylene or polypropylene. The dimensions of the orifice in the actuator plays a crucial role, along with the propellant vapour pressure, in determining the shape and speed of the emitted aerosol plume.

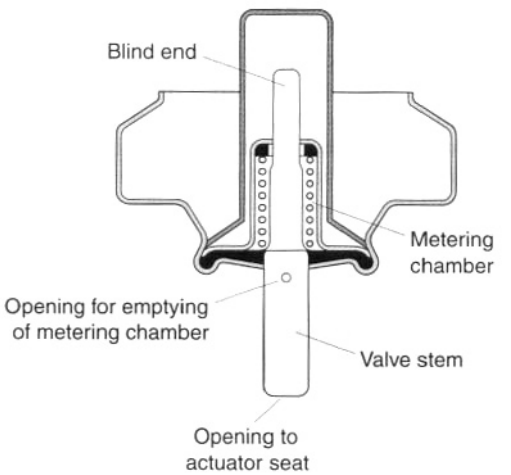


Fig. 31.3 The metering valve. (Reproduced with permission from Morén 1981.)

Formulating metered-dose inhalers

Pressurized aerosols may be formulated as either solutions or suspensions of drug in the liquefied propellant. Solution preparations are two-phase systems. However, the propellants are poor solvents for most drugs. Cosolvents such as ethanol or isopropanol may be used, although their low volatility retards propellant evaporation. In practice, pressurized inhaler formulations have, until recently, been almost exclusively suspensions. These three-phase systems are harder to formulate and all the problems of conventional suspension formulation, such as caking, agglomeration, particle growth etc. must be considered. Careful consideration must be given to the particle size of the solid (usually micronized to between 2 and 5 μm), valve clogging, moisture content, the solubility of active compound in propellant (a salt may be desirable), the relative density of propellant and drug, and the use of surfactants as suspending agents, e.g. lecithin, oleic acid and sorbitan trioleate (usually included at concentrations between 0.1 and 2.0% w/w). These surfactants are very poorly soluble (<< 0.02% w/w) in HFAs, and so either ethanol must be used as a cosolvent or alternative surfactants such as fluorinated polymers must be developed (Byron et al 1994). Recently solution formulations of beclomethasone dipropionate have been marketed. Evaporation of HFA propellant following actuation of these formulations results in smaller particle sizes than with conventional suspension formulations of the same drug, with consequent changes in its pulmonary distribution and bioavailability.

Filling metered-dose inhaler canisters

Canisters are filled by liquefying the propellant at reduced temperature or elevated pressure.

In ***cold filling***, active compound, excipients and propellant are chilled and filled at about −30°C. Additional propellant is then added at the same temperature and the canister sealed with the valve. In ***pressure filling***, a drug/propellant (CFC-11) concentrate is produced and filled at effectively room temperature and pressure (in fact, usually slightly chilled to below 20°C). The valve is crimped on to the canister and additional propellant (e.g. CFC-12) is filled at elevated pressure through the valve, in a process known as gassing. Pressure filling is most frequently employed for inhalation aerosols. However, no ozone-sparing replacement propellant has the properties (high boiling point: 23.7°C) of CFC-11, which is a major problem for the pharmaceutical industry.

Once filled, the canisters are leak tested by placing them in a water bath at elevated temperature, usually 50–60°C. Following storage to allow equilibration of the formulation and valve components, the containers are weighed to check for further leakage, prior to spray testing and insertion into actuators.

Advantages and disadvantages of metered-dose inhalers

The major advantages of MDIs are their portability, low cost and disposability. Many doses (up to 200) are stored in the small canister and dose delivery is reproducible. The inert conditions created by the propellant vapour, together with the hermetically sealed container, protects drugs from oxidative degradation and microbiological contamination. However, MDIs have disadvantages. They are inefficient at drug delivery. On actuation, the first propellant droplets exit at a high velocity, which may exceed 30 m/s. Consequently, much of the drug is lost through impaction of these droplets in the oropharyngeal areas. The mean emitted droplet size typically exceeds 40 μm, and propellants may not evaporate sufficiently rapidly for their size to decrease to that suitable for deep lung deposition. Vaporization of the droplets is hindered by the low volatility of CFC-11, which is present in concentrations of at least 25% in most CFC-based formulations. Evaporation, such that the aerodynamic diameter of the particles is close to that of the original micronized drug, may not occur until 5 seconds after actuation.

An additional problem with MDIs, which is beyond the control of the formulator and manufacturer, is their incorrect use by patients. Reported problems include:

- Failure to remove the protective cap covering the mouthpiece, the inhaler being used inverted;
- Failure to shake the canister;
- Failure to inhale slowly and deeply;
- Inadequate breath-holding;
- Poor inhalation/actuation synchronization.

Correct use by patients is vital for effective drug deposition and action. Ideally, the MDI should be actuated during the course of slow, deep inhalation, followed by a period of breath-holding. Many patients find this difficult, especially children and the elderly. The misuse of MDIs through poor inhalation/actuation coordination can be significantly reduced with appropriate instruction and counselling. However, it should be noted that even using the correct inhalation technique only 10–20% of the stated emitted dose is delivered to the site of action.

Spacers and breath-actuated metered-dose inhalers

Some of the disadvantages of MDIs, namely inhalation/actuation coordination and the premature deposition of large propellant droplets high in the airways, can be overcome by using extension devices or 'spacers' positioned between the MDI and the patient (Fig. 31.4). The dose from an MDI is discharged directly into the reservoir prior to inhalation. This reduces the initial droplet velocity, permits efficient propellant evaporation and removes the need for actuation/inhalation coordination. The disadvantage of spacers is that they may be cumbersome, e.g. Fisonair (Rhône-Poulenc Rorer), Nebuhaler (AstraZeneca), and Volumatic (Glaxo SmithKline). Alternatively, extension tubes may be built into the design of the MDI itself, e.g. Syncroner (Rhône-Poulenc Rorer) and Spacer Inhalers (AstraZeneca). The Autohaler (3M) is an MDI with

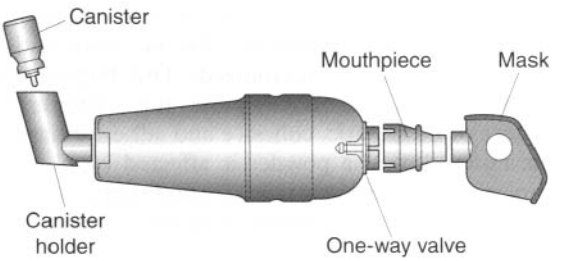


Fig. 31.4 The Nebuhaler spacer device, fitted with a face mask for use by a child. Courtesy of AstraZeneca.

an inspiratory demand valve. This breath-actuated device overcomes the coordination problems of a conventional MDI without adding bulk to the device. However, a substantial inspiratory flow rate is required for its operation.

Dry powder inhalers

In dry powder inhaler (DPI) systems, drug is inhaled as cloud of fine particles. The drug is either preloaded in an inhalation device or filled into hard gelatin capsules or foil blister discs which are loaded into a device prior to use. DPIs have several advantages over MDIs. DPI formulations are propellant free and do not contain any excipient, other than a carrier (see below) – which is almost invariably lactose. They are breath actuated, avoiding the problems of inhalation/actuation coordination encountered with MDIs, and consequently they are particularly useful for young children. DPIs can also deliver larger drug doses than MDIs, which are limited by the volume of the metering valve and the maximum suspension concentration that can be employed without causing valve clogging. However, DPIs have several disadvantages. Liberation of powders from the device and the deaggregation of particles are limited by the patient's ability to inhale, which in the case of respiratory disease may be impaired. An increase in turbulent air flow created by an increase in inhaled air velocity increases the deaggregation of the emerging particles, but also increases the potential for inertial impaction in the upper airways and throat, and so a compromise has to be found. Further, DPIs are exposed to ambient atmospheric conditions, which may reduce formulation stability. For instance, elevated humidity may cause powders to clump. Finally, DPIs are generally less efficient at drug delivery than MDIs, such that twice the dose is usually required for delivery from a DPI than from the equivalent MDI (Melchor et al 1993).

Formulating dry powder inhalers

To produce particles of a suitable size (preferably less than 5 μm), drug powders for use in inhalation systems are usually micronized. The high-energy powders produced have poor flow properties because of their static, cohesive and adhesive nature. The flowability of a powder is affected by physical properties, including particle size and shape, density, surface roughness, hardness, moisture content and bulk density.

To improve their flow properties, poorly flowing drug particles are generally mixed with larger 'carrier' particles (usually 30–60 μm) of an inert excipient, usually lactose. This not only improves liberation of the drug from the inhalation device by improving powder flow, but also improves the uniformity of capsule or device filling. Once liberated from the device, the turbulent air flow generated within the inhalation device should be sufficient for the deaggregation of the drug/carrier aggregates. The larger carrier particles impact in the throat, whereas smaller drug particles are carried in the inhaled air deeper into the respiratory tract.

The success of DPI formulations depends on the adhesion of drug and carrier during mixing and filling of devices or hard gelatin capsules, followed by the ability of the drug to desorb from the carrier during inhalation such that free drug is available to penetrate to the peripheral airways. Adhesion and desorption will depend on the morphology of the particle surfaces and surface energies, which may be influenced by the chemical nature of the materials involved and the nature of powder processing. The performance of DPI systems is thus strongly dependent on formulation factors, and also on the construction of the delivery device and the inhalation technique.

Unit-dose devices with drug in hard gelatin capsules

The first DPI device developed was the Spinhaler (Rhône-Poulenc Rorer) for the delivery of sodium cromoglycate (Fig. 31.5). Each dose, contained in a hard gelatin capsule, is loaded individually into the device. The capsule, placed in a loose-fitting rotor, is pierced by two metal needles an either side of the capsule. Inhaled air flow though the device causes a

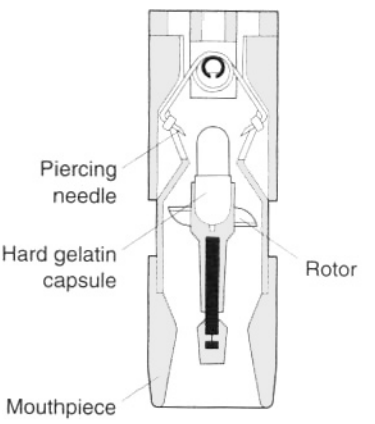


Fig. 31.5 The Spinhaler. (Modified from Bell et al 1971, with the permission of the American Pharmaceutical Association.)

turbovibratory air pattern as the rotor rotates rapidly, resulting in the powder being dispersed to the capsule walls and out through the perforations into the air. A minimum air flow rate of 35–40 L/min through the device is required to produce adequate vibrations by the rotor. The occurrence of lactose intolerance and local irritation, coughing and bronchoconstriction caused by the inhalation of large amounts of lactose has led to the development of an aggregated, carrier-free sodium cromoglycate capsule formulation for use in the Spinhaler.

Another unit-dose DPI is the Rotahaler (Glaxo SmithKline), which is a simple two-piece device (Fig. 31.6). The gelatin capsule is inserted into an orifice at the rear of the device and when the two sections are rotated a fin on the inner barrel pulls the two halves of the capsule apart. During inhalation, the freed half of the capsule spins, dispersing its contents, which are inhaled through the mouthpiece. The resistance to air flow is lower than that of the Spinhaler and therefore a lower inspiratory velocity is required.

Other hard gelatin capsule-based devices, working on similar principles, are available for the delivery of drug/carrier mixes. These include the Aerohaler (Boehringer Ingelheim) and the Cyclohaler (Du Pont).

Multidose devices with drug in foil blisters

The main disadvantage of hard gelatin capsule-based devices, namely the individual loading of each dose, was overcome with the development of the Diskhaler (Glaxo SmithKline). In this system, drug is mixed with a coarse lactose carrier and filled into an aluminium foil blister disc which is loaded, by the patient, into the device on a support wheel (Fig. 31.7). Each disc contains four or eight doses of drug and the blisters are pierced with a needle as a result of mechanical leverage on the lid. Air flow through the blister causes the powder to disperse as the patient inhales through the mouthpiece. The foil blisters are numbered, so that the patient knows the number of doses remaining.

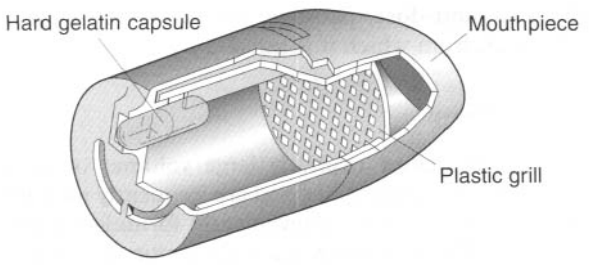


Fig. 31.6 The Rotahaler (Modified with permission from Kjellman 1981.)

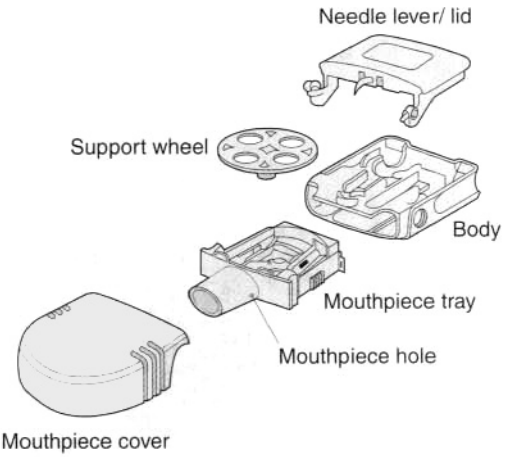


Fig. 31.7 The Diskhaler (Reproduced with permission from Sumby et al 1993.)

Multidose devices with drug preloaded in inhaler

The evolution of the Diskhaler led to the production of the Accuhaler or Diskus Inhaler (Glaxo SmithKline), in which drug/carrier mix is preloaded into the device in foil-covered blister pockets containing 60 doses (Fig. 31.8). The foil lid is peeled off the drug-containing pockets as each dose is advanced, with the blisters and lids being wound up separately within the device, which is discarded at the end of operation. As each dose is packaged separately and only momentarily exposed to ambient conditions prior to inhalation, the Diskhaler and Accuhaler are relatively insensitive to humidity compared to hard gelatin capsule-based systems.

An alternative approach is a ***reservoir*** type of device, in which a dose is accurately measured and delivered from a drug reservoir. In the Clickhaler DPI (Innovata Biomed), a drug blend is stored in a reservoir. Metering cups are filled by gravity from this reservoir and delivered to an inhalation passage, from which it is inhaled. The device is capable of holding up to 200 doses and incorporates a dose counter, which informs patients when the device, which is discarded after use, is nearly empty.

The Turbohaler (AstraZeneca), has overcome the need for both a carrier and the loading of individual doses (Fig. 31.9). The device contains a large number of doses (up to 200) of undiluted, loosely aggregated micronized drug, which is stored in a reservoir from which it flows on to a rotating disc in

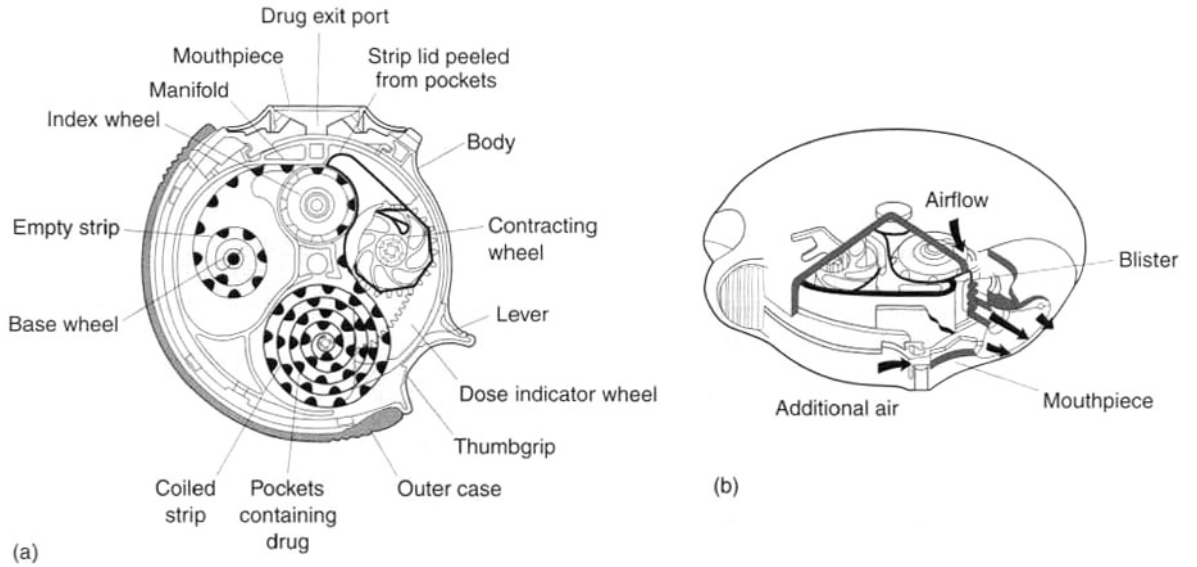


Fig. 31.8 The Accuhaler/Diskus Inhaler, showing (a) a schematic diagram and (b) a cross-sectional representation of the device. (Reproduced with permission from Prime et al 1996.)

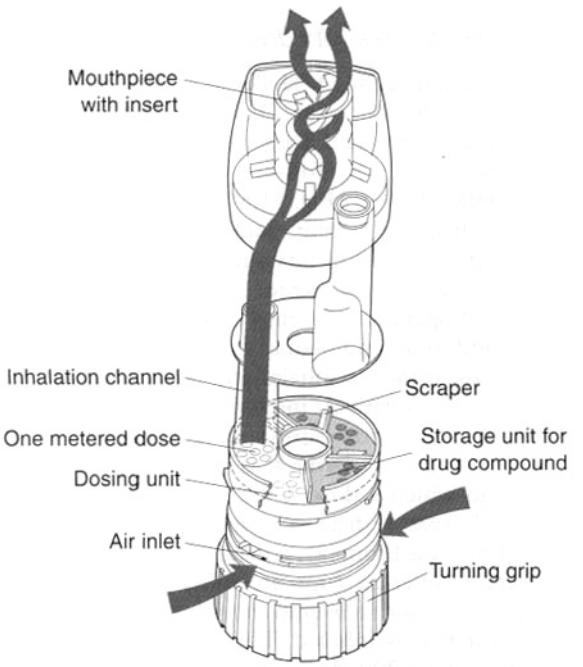


Fig. 31.9 The Turbohaler. (Reproduced with permission from Wetterlin 1987.)

the dosing unit. The fine holes in the disc are filled and the excess drug is removed by scrapers. As the rotating disc is turned, by moving a turning grip back and forth, one metered dose is presented to the inhalation channel, and this is inhaled by the patient, with the turbulent air flow created within the device breaking up any drug aggregates. A dose indicator is incorporated. The Turbohaler requires a higher inspiratory effort than the Diskhaler, owing to its higher internal resistance, and is more sensitive to humidity if not closed quickly after each use.

Non-breath actuated devices

Devices are currently under development which reduce or eliminate the reliance on the patient's inspiratory effort to disperse the drug (Rubsamen 1997). Such inspiratory effort may be affected by the patient's age and/or clinical condition. For instance, a device that uses a battery-powered impeller to deaggregate the drug powder is being developed. The device is breath actuated, but deaggregation is independent of the patient's inspiratory flow rate. Inhale Therapeutic Systems have produced a device in which compressed air is used to disperse drug from a unit-dose package into a large holding chamber, from which it is inhaled by the patient.

Nebulizers

Nebulizers deliver relatively large volumes of drug solutions and suspensions and are frequently used for drugs that cannot be conveniently formulated into MDIs or DPIs, or where the therapeutic dose is too large for delivery with these alternative systems. Nebulizers also have the advantage over metered-

dose and dry powder systems in that drug may be inhaled during normal tidal breathing through a mouthpiece or face-mask, and thus they are useful for patients such as children, the elderly and patients with arthritis, who experience difficulties with MDIs.

There are two categories of commercially available nebulizer: jet and ultrasonic.

Jet nebulizers

Jet nebulizers (also called air-jet or air-blast nebulizers) use compressed gas (air or oxygen) from a compressed gas cylinder, hospital air-line or electrical compressor to convert a liquid (usually an aqueous solution) into a spray. The jet of high-velocity gas is passed either tangentially or coaxially through a narrow Venturi nozzle, typically 0.3–0.7 mm in diameter. An area of negative pressure, where the air jet emerges, causes liquid to be drawn up a feed tube from a fluid reservoir by the Bernoulli effect (Fig. 31.10). Liquid emerges as fine filaments, which collapse into droplets owing to surface tension. A proportion of the resultant (primary) aerosol leaves the nebulizer directly; the remaining, large, non-respirable droplets impact on baffles or the walls of the nebulizer chamber and are recycled into the reservoir fluid.

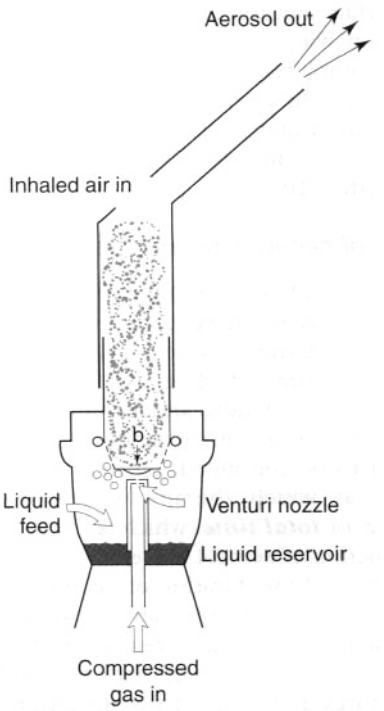


Fig. 31.10 Schematic diagram of a jet nebulizer. Compressed gas passes through a Venturi nozzle, where an area of negative pressure is created. Liquid is drawn up a feed tube and is fragmented into droplets. Large droplets impact on baffles (b), and small droplets are carried away in the inhaled airstream. (Reproduced with permission from Newman 1989.)

Nebulizers are operated continuously, and because the inspiratory phase of breathing constitutes approximately one-third of the breathing cycle a large proportion of the emitted aerosol is not inhaled but is released into the environment. Open-vent nebulizers, incorporating inhalation and exhalation valves, e.g. the Pari LC nebulizer (Pari) have recently been developed in which the patient's own breath boosts nebulizer performance, with aerosol production matching the patient's tidal volume and greatly enhancing drug delivery. On exhalation, the aerosol being produced is generated only from the compressor gas source, thereby minimizing drug wastage.

The rate of gas flow driving atomization is the major determinant of the aerosol droplet size and rate of drug delivery for jet nebulizers: for instance, there may be up to a 50% reduction in the mass median aerodynamic diameter (MMAD, see below) when the flow rate is increased from 4 to 8 L/min, with a linear increase in the proportion of droplets less than 5 μm (Clay et al 1983).

Ultrasonic nebulizers

In ultrasonic nebulizers the energy necessary to atomize liquids comes from a piezoelectric crystal vibrating at high frequency. At sufficiently high ultrasonic intensities a fountain of liquid is formed in the nebulizer chamber. Large droplets are emitted from the apex and a 'fog' of small droplets is emitted from the lower part (Fig. 31.11). Some models have a fan to blow the respirable droplets out of the device, whereas in others the aerosol only becomes available to the patient during inhalation.

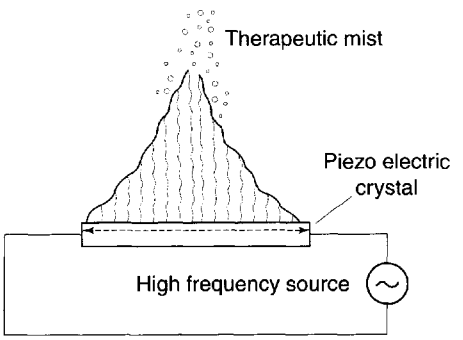


Fig. 31.11 Schematic diagram of an ultrasonic nebulizer. (Reproduced with permission from Atkins et al 1992.)

Formulating nebulizer fluids

Nebulizer fluids are formulated in water, occasionally with the addition of cosolvents such as ethanol or propylene glycol, and with the addition of surfactants for suspension formulations. Because hypoosmotic and hyperosmotic solutions may cause bronchoconstriction, as may high hydrogen ion concentrations, iso-osmotic solutions of pH greater than 5 are usually employed (Snell 1990). Stabilizers such as antioxidants and preservatives may also be included, although these may also cause bronchospasm, and for this reason sulphites in particular are generally avoided as antioxidants in such formulations. Although chemically preserved multidose preparations are commercially available, nebulizer formulations are generally presented as sterile, isotonic unit doses (usually 1–2.5 mL) without a preservative.

Whilst most nebulizer formulations are solutions, suspensions of micronized drug are also available for delivery from nebulizers. In general suspensions are poorly delivered from ultrasonic nebulizers, whereas with jet nebulizers the efficiency of drug delivery increases as the size of suspended drug is decreased, with little or no release of particles when they exceed the droplet size of the nebulized aerosol.

As the formulation of fluids for delivery by nebulizers is relatively simple, these devices are frequently the first to be employed when investigating the delivery of new entities to the human lung. Recently, they have been used for the delivery of peptides and liposomes. In general, ultrasonic nebulizers have not been successful for delivering either peptides or liposomes, because of denaturation resulting from the elevated temperatures produced. Consequently, ultrasonic nebulizers are expressly excluded for the delivery of recombinant human deoxyribonuclease in the management of cystic fibrosis. Jet nebulizers have been successfully used to deliver some peptide and liposome formulations, although the shearing forces that occur in the nebulizer may produce time-dependent damage to some materials (Niven and Brain 1994).

Physicochemical properties of nebulizer fluids

The viscosity and surface tension of a liquid being nebulized may affect the output of nebulizers, as energy is required to overcome viscous forces and to create a new surface. However, the size-selectivity of the nebulizer design and dimensions, with more than 99% of the primary aerosol mass being recycled into the reservoir liquid, means that changes in the size distribution of the primary aerosol resulting from changes in the properties of the solution being atomized may not always be reflected in the size distribution of the emitted aerosol. In general, the size of aerosol droplets is inversely proportional to viscosity for jet nebulizers and directly proportional to viscosity for ultrasonic nebulisers (McCallion et al 1995), with more viscous solutions requiring longer to nebulize to dryness and leaving larger residual volumes in the nebulizer following atomization. Surface tension effects are more complex, but usually a decrease in surface tension is associated with a reduction in mean aerosol size.

Temperature effects during nebulization

The aerosol output from a jet nebulizer comprises drug solution and solvent vapour, which saturates the outgoing air. This causes solute concentration to increase with time and results in a rapid decrease in the temperature of the liquid being nebulized by approximately 10–15°C.

This temperature decrease may be important clinically, as some asthma sufferers experience bronchoconstriction on inhalation of cold solutions. Further, the cooling effect within the reservoir fluid will reduce drug solubility and result in increased liquid surface tension and viscosity. Precipitation is uncommon with bronchodilators, which have high aqueous solubility, but problems may arise with less soluble drugs. In such instances the use of an ultrasonic nebulizer may be appropriate, as the operation of such devices increases solution temperature by approximately 10–15°C.

Duration of nebulization and 'dead volume'

Clinically, liquids may be nebulized for a specified period of time, or more commonly, they may be nebulized to 'dryness', which may be interpreted as ***sputtering time***, which is the time when air is drawn up the feed tube and nebulization becomes erratic, although agitation of the nebulizer permits treatment to be continued; ***clinical time***, which is the time at which therapy is ceased following sputtering; or ***total time***, which is the time at which the production of aerosol ceases.

Regardless of the duration of nebulization, not all the fluid in the nebuliser can be atomized. Some liquid, usually about 1 mL, remains as the 'dead' or 'residual' volume, associated with the baffles, internal structures and walls of the nebulizer. The proportion of drug retained as 'dead' volume is more marked for smaller fill volumes, hence for a 2 mL fill volume, approximately 50% of fluid will remain associated with the nebulizer and be unavailable for delivery to the patient. This reduces to approxi-

mately 25% with a 4 mL fill volume, although there is a commensurate increase in the time necessary to nebulize to dryness.

Variability between nebulizers

Many different models of nebulizer and compressor are commercially available, and the size of aerosols produced and the dose delivered can vary enormously. For instance, in a study of 18 different commercially available jet nebulizers, operated according to the manufacturers' guidelines, aerosols were produced with MMADs ranging from 0.9 to 7.2 μm (Waldrep et al 1994). Variability may not only exist between different nebulizers but also between individual nebulizers of the same type, and repeated use of a single nebulizer may cause variability due to baffle wear and non-uniformity of assembly. Nebulizers, unlike the DPI and MDI devices, are not manufactured by the producers of nebulizer solutions and suspension. The choice of nebulizer employed for their delivery is thus usually beyond the influence of the pharmaceutical manufacturer.

METHODS OF AEROSOL SIZE ANALYSIS

The regional distribution of aerosols in the airways can be measured directly using gamma scintigraphy, by radiolabelling droplets or particles, usually with the short half-life gamma emitter technetium-99m(^{99m}Tc). However, more commonly in vitro measurements of aerosol size are used to predict clinical performance. The principal methods that have been employed for size characterization of aerosols are microscopy, laser diffraction and cascade impaction.

Optical methods of measuring the physical size of deposited aerosols using microscopy are laborious and do not give an indication of their likely deposition within the humid airways while being carried in an airstream. With methods of analysis based on laser Fraunhofer diffraction, aerosolized droplets or particles are sized as they traverse a laser beam to give a volume median diameter. Again the aerodynamic properties of an aerosol are not being measured. In addition, spraying droplets into a beam exposes them to ambient conditions of temperature and humidity, which may result in solvent evaporation.

Cascade impactors and impingers

Cascade impactors comprise a series of progressively finer jets and collection plates, allowing fractionation of aerosols according to their MMAD as the aerosol is drawn through the device at a known flow rate. Traditional cascade impactors are constructed from metal. The most commonly used comprises eight stages, with metal collection plates followed by a terminal filter. Multistage liquid impingers, working on the same principle, are constructed from glass or glass and metal and have three, four or five stages, with wet sintered glass collection plates followed by a terminal filter. Large dense particles will deposit higher in the impactor, whereas smaller, less dense particles will follow the air flow and only deposit when they have been given sufficient momentum as they are accelerated through the finer jets lower in the impactor (Fig. 31.12). The first stage of the impactor is usually preceded by a 90° bend of metal or glass to mimic the human throat. The cut-off diameters for each stage at a particular air-flow rate can be determined using monodisperse aerosols or calculated using calibration curves. When determining the size of an aerosol, cumulative percentage undersize plots of deposited aerosol on each stage are plotted against the cut-off diameter for that stage to allow calculation of the MMAD.

The five-stage liquid impinger (MSLI) (Fig. 31.13), with an appropriate induction port and mouthpiece adapter, is used to determine the aerodynamic size of DPIs (USP and EP), MDIs and nebulizers (EP). The MSLI may be operated at a flow rate between 30 and 100 L/min. At 60 L/min (i.e. 1 L/s) the effective cut-off diameters of stages 1, 2, 3 and 4 are 13.0, 6.8, 3.1 and 1.7 μm, respectively. The fifth stage comprises an integral filter which captures particles smaller than 1.7 μm. When testing DPIs to USP requirements, an airflow rate (Q) calculated to produce a pressure drop of 4.0 kPa over the inhaler is employed. If this exceeds 100 L/min, then 100 L/min is used. The cut-off diameters of each stage at flow rate (Q) can be calculated from:

$$D_{50'Q} = D_{50'Qn}(Q_n/Q)^{\frac{1}{2}} \qquad (31.7)$$

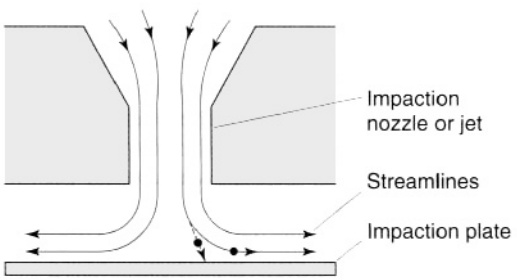


Fig. 31.12 Illustration of aerodynamic particle size separation by an impactor stage. (Reproduced with permission from Jaegfeldt et al 1987.)

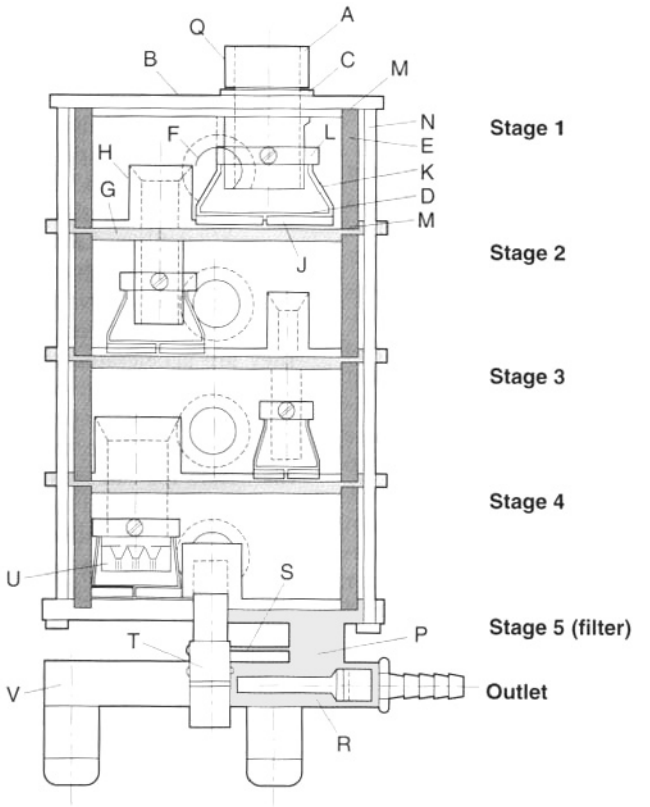


Fig. 31.13 The multistage liquid impinger. Courtesy of AstraZeneca.

where $D_{50'Q}$ is the cut-off diameter at the flow rate Q and n refers to the nominal cut-off values determined when Q_n is 60 L/min (values given above).

The use of cascade impaction methods to determine the size of aerosols has a number of disadvantages. The high flow rates employed (typically 28.3–60 L/min) result in rapid solvent evaporation, and droplets may be re-entrained in the airstream whereas particles may 'bounce off' metal collection plates, although this latter effect may be reduced by coating the collection surface, for instance with a silicone fluid or glycerol. These effects can result in a significant decrease in the measured aerosol size. Also, these measuring devices are operated at a constant air-flow rate. However, the dispersion of dry powder formulations and the deposition profile of inhaled aerosols will very considerably with flow rate. To overcome the limitations of measurement at a single flow rate the Electronic Lung (The Technology Partnership) has been developed, which uses a computer-controlled piston to draw air through the inhaler and into an impaction sizer, following a predetermined inhalation profile (Brindley et al 1994).

Cascade impactor methods are invasive, laborious and time-consuming, but necessary to derive information about median aerosol size and the polydispersity of the aerosol. To ensure that inhalation products are likely to be clinically effective, quality control measurements usually involve measurement of the emitted dose and the 'fine particle fraction', (that fraction of the emitted dose less than a stated size, often 5 or 6.4 μm), which are combined to give a 'useful' or 'respirable' dose or mass (Ganderton, 1995). For routine analysis, a simplified two-stage (twin) impinger is frequently employed (Fig. 31.14). Aerosol collected in the throat and the upper stage (stage 1) is considered 'non-respirable', whereas that collected in the lower stage (stage 2) is considered 'respirable'. For this glass device, the cut-off diameter for stage 2 is 6.4 μm, i.e. aerosols collected in this stage have an aerodynamic diameter less than 6.4 μm and are for this measurement technique considered 'respirable'. This twin impinger is included in the BP as a method for determining the emitted dose from MDIs and DPIs.

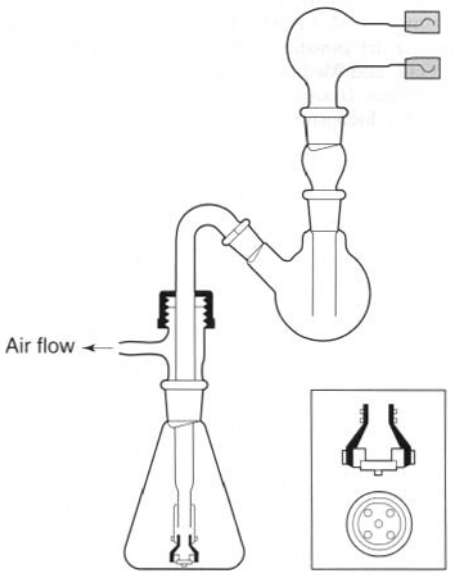


Fig. 31.14 The two-stage impinger. (Reproduced with permission from Hallworth and Westmoreland 1987.)

REFERENCES

Atkins, P.J, Barker, N.P., Mathisen, D. (1992) The design and development of inhalation drug delivery systems. In: *Pharmaceutical Inhalation Aerosol Technology*. (Ed. A.S.J. Hickey) Marcel Dekker, New York, pp155–185.

Bell, J.H., Hartley, P.S. and Cox, J.S.G. (1971) Dry powder aerosols I: a new powder inhalation device. *J. Pharm Sci.*, **60**, 1559–1564.

Brindley, A., Sumby, B.S., Smith, I.J. (1994) The characterisation of inhalation devices by an inhalation simulator: the electronic lung. *J Aerosol Med.*, **7**, 197–200.

Byron, P.R., Miller, N.C., Blondino, F.E., Visich, J.E., Ward, G.H. (1994). Some aspects of alternative propellant solvency. *Respir. Drug Del. IV*, 231 242.

Clay, M.M., Pavia, D., Newman, S.P., Lennard-Jones, T., Clarke, S.W. (1983) Assessment of jet nebulisers for lung aerosol therapy. *Lancet*, **ii**, 592 594.

Ganderton, D. (1995) In-vitro testing of inhalation products. *Drug Delivery to the Lungs VI*. The Aerosol Society, Bristol, pp 85–88.

Hallworth, G.W., Westmoreland, D.G. (1987) The twin impinger: a simple device for assessing the delivery of drugs from metered dose pressurised aerosol inhalers. *J. Pharm. Pharmacol.*, **39**, 966 972.

Jaegfeldt, H., Andersson, J.A.R., Trofast, E., Wetterlin, K.I.I.. (1987) Particle size distribution from different modifications of Turbohaler. In: *A New Concept In Inhalation Therapy*. (Eds. S.P. Newman, F. Morén and G.K. Crompton). Medicom Europe, Bussum, pp 90 99.

Kjellman, N-I.M. (1981) Letter to the Editor *Allergy*, **36**, 437 438.

McCallion, O.N.M., Taylor, K.M.G., Thomas, M., Taylor A.J. (1995) Nebulization of fluids of different physicochemical properties with air-jet and ultrasonic nebulizers. *Pharm. Res.*, **12**, 1682 1688.

Melchor, R., Biddiscombe, M.F., Mak, V.H.F., Short, M.D., Spiro, S.G. (1993) Lung deposition patterns of directly labelled salbutamol in normal subjects and in patients with reversible airflow obstruction. *Thorax*, **48**, 506-511.

Morén, F. (1981) Pressurised aerosols for oral inhalation. *Int. J. Pharm.*, **8**, 1–10.

Newman, S.P. (1989) *Nebuliser Therapy: Scientific and Technical Aspects*. AB Draco, Lund.

Niven, R.W. and Brain, J.D. (1994) Some functional aspects of air-jet nebulizers. *Int. J. Pharm.*, **104**, 73–85.

Prime, D., Slater, A.L., Haywood, P.A., Smith, I.J. (1996) Assessing dose delivery from the Flixotide Diskus Inhaler a multidose powder inhaler. *Pharm. Tech. Europe*, **8 (3)**, 23–34.

Pritchard, J.N. (1987) Particle growth in the airways and the influence of airflow. In: *A New Concept In Inhalation Therapy*. (Eds. S.P. Newman, F. Morén and G.K. Crompton). Medicom Europe, Bussum, pp 3–24.

Rubsamen, R. (1997) Novel aerosol peptide drug delivery systems. In: *Inhalation Delivery of Therapeutic Peptides and Proteins*. (Eds. A.L. Adjei and P.K. Gupta). Marcel Dekker, New York, pp 703–731.

Snell, N.J.C. (1990) Adverse reactions to inhaled drugs. *Respir. Med.*, **84**, 345–348.

Sumby, B.S., Churcher, K.M., Smith, I.J., Grant, A.C., Truman, K.G., Marriott, R.J., Booth, S.J. (1993) Dose reliability of the Serevent Diskhaler system. *Pharm. Tech. Int.*, **6 (5)**, 20–27.

Waldrep, J.C., Keyhani, K., Black, M., Knight, V. (1994) Operating characteristics of 18 different continuous flow jet nebulizers with beclomethasone dipropionate liposome aerosol. *Chest*, **105**, 106–110.

Wetterlin, K.I.L. (1987) Design and function of Turbohaler. In: *A New Concept In Inhalation Therapy*. (Eds. S.P. Newman, F. Morén and G.K. Crompton). Medicom Europe, Bussum, pp 85–89.

Wilson, C.G. and Washington N. (1989) *Physiological Pharmaceutics: Biological Barriers to Drug Absorption*. Ellis Horwood, Chichester, p157.

BIBLIOGRAPHY

Adjei, A.L. and Gupta, P.K. (eds) (1997) *Inhalation Delivery of Therapeutic Peptides and Proteins*. Marcel Dekker, New York.

Farr, S.J, Kellaway, I.W. and Taylor, G. (1990). Drug delivery to the respiratory tract. In: *Routes of Drug Administration*. (Eds. A.T. Florence and E. Salole). Wright, London, pp 48–77.

Hickey, A.S.J. (ed) (1992) *Pharmaceutical Inhalation Aerosol Technology*. Marcel Dekker, New York.

McCallion, O.N.M. and Taylor, K.M.G. (1999) Ultrasonic nebulizers. In: *Encyclopaedia of Pharmaceutical Technology*. (Eds J. Swarbrick and J.C. Boylan). Marcel Dekker, New York, pp339–352.

McCallion, O.N.M., Taylor K.M.G., Bridges, P.A., Thomas, M., Taylor A.J. (1996) Jet nebulisers for pulmonary drug delivery. *Int. J. Pharm.*, **130**, 1–11.

Morén, F. (1981) Pressurised aerosols for oral inhalation. *Int. J. Pharm.*, **8**, 1–10.

Sciarra, J.J. (1996) Pharmaceutical Aerosols. In: *Modern Pharmaceutics*. 3rd edn., (Eds. G.S. Banker and C.T. Rhodes), Marcel Dekker, New York, pp 547–574.

Sciarra, J.J. and Cutie A.J. (1990) Aerosols. In: *Remington's Pharmaceutical Sciences*. 18th ed. (Eds. A.R. Gennaro et al). Mack Publishing Company, Easton, pp 1694–1712.

Timsina, M.P., Martin G.P., Marriott, C., Ganderton, D., Yianneskis, M. (1994) Drug delivery to the respiratory tract using dry powder inhalers. *Int. J. Pharm.*, **101**, 1–13.

Wilson, C.G. and Washington N. (1989) *Physiological Pharmaceutics: Biological Barriers to Drug Absorption*. Ellis Horwood, Chichester.

32
Nasal drug delivery

Peter Taylor

CHAPTER CONTENTS

INTRODUCTION

The nasal cavity has been considered as a route of drug administration for many decades, often for topical therapies such as decongestants. Recently, however, there has been a great deal of research investigating the nose as a route for systemic therapies, especially for peptides and proteins (see Tables 32.1 to 32.3 for examples of topical nasal preparations, systemic preparations on the market, and systemic preparations under investigation).

There are many potential benefits to nasal administration: it is convenient, there is a useful area for absorbing drugs, and there is a good systemic blood supply – but, as with many other routes of delivery, there are also factors working against efficient drug absorption. The aim of this chapter is to review the relevant features of nasal anatomy and physiology, outline the various factors that influence drug absorption and, finally, discuss the various formulation strategies and delivery systems that can be used for nasal drug delivery.

THE ANATOMY AND PHYSIOLOGY OF THE NOSE

The outermost part of the nose is the nasal vestibule, which runs for about 15 mm from the nostrils to the nasal valve (Fig. 32.1). Behind the nasal valve is the nasal cavity, with a length of about 60 mm and a volume of 20 mL, and this passes into the nasopharynx. The nasal cavity is divided vertically for most of its length by the nasal septum, and each wall of the cavity contains three folds or indentations known as the nasal turbinates (or conchae). This folding means that the nasal cavity has a relatively large surface area for its volume – approximately 160 cm^2 (Chien 1992, Lee and Baldwin 1992, Mygind and Dahl 1998).

Table 32.1 Examples of topical nasal preparations (taken from the British National Formulary 39)

Drug	Drug class	Use	Delivery system
Levocabastine	Antihistamine	Allergic rhinitis	Aqueous spray (pump)
Beclomethasone dipropionate	Corticosteroid	Allergic and vasomotor rhinitis	Metered spray (pressurized) of aqueous suspension
Sodium cromoglycate	Cromoglycate	Allergic rhinitis	Aqueous spray (pump)
Ephedrine HCl	Sympathomimetic	Decongestant	Drops
Chlorhexidine HCl	Antimicrobial	Staphylococcal infections (nostrils)	Cream

Table 32.2 Examples of marketed systemic nasal preparations (taken from the British National Formulary 39 and Behl et al 1998a)

Drug	Drug class	Use	Delivery system
Desmopressin	Pituitary hormone	Diabetes insipidus	Metered spray
Buserelin	Gonadorelin analogue	Prostate cancer, endometriosis and others	Metered spray (pump)
Nafarelin	Gonadorelin analogue	Prostate cancer, endometriosis and others	Metered spray
Sumatriptin	Serotonin agonist	Migraine	Unit dose spray
Dihydroergotamine mesylate	Ergot alkaloid	Migraine	4-application spray
Nicotine	–	Cessation of smoking	Metered spray
Salmon calcitonin	Calcium regulator	Postmenopausal osteoporosis	Metered spray

Table 32.3 Examples of drugs under investigation for nasal delivery (taken from Behl et al 1998a)

Drug	Use	Stage of investigation
Cyanocobalamin (Vitamin B_{12}) Gel	Vitamin B_{12} deficiency	NDA submitted (USA)
17-β-oestradiol	Relief of menopausal symptoms	Human studies
Glucagon	Treatment of hypoglycaemia	Human studies
Interferon	Antiviral agent	Human studies
Insulin	Treatment of diabetes	Human studies
Metoclopramide HCl	Antiemetic	Human studies
Vaccines	Influenza, measles, polio etc.	Human studies

The normal functioning of the nose is closely related to its anatomy, for not only is it a sensory organ, it also conditions inspired air, heating and humidifying it before it reaches the lungs. Air passes through the constriction of the nasal valve with a high linear velocity, but the sharp change in direction of the flow into the nasal cavity and the presence of the turbinates (see Fig. 32.1) cause turbulent flow, bringing the air into close contact with the nasal lining. The large surface area of the nasal cavity and the rich underlying vasculature cause rapid heat and moisture transfer to the air; these are also factors that predispose the nasal cavity to being a good site for drug absorption.

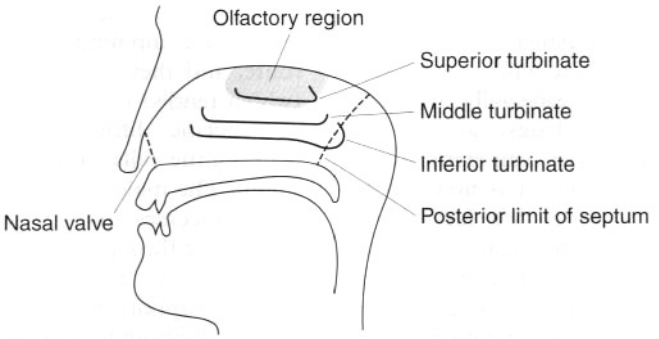


Fig. 32.1 Cross-section of human nasal cavity, showing major structures relevant to nasal drug administration.

Nasal mucosa and mucociliary clearance

The anterior part of the nose, from the nasal vestibule to the turbinates, consists of squamous epithelium. The upper part of the cavity, accounting for about 5% of the total area, is olfactory membrane. The latter contains the sensory olfactory cells as well as serous and mucosal cells; it is so placed because a large proportion of inspired air passes over this region. Most of the nasal cavity, however, is lined with mucous membrane containing a mixture of columnar cells, goblet cells and basal cells. The columnar cells in the forward third of the epithelium are non-ciliated; the remainder are covered with cilia. Cilia are hair-like projections on the exposed surface of epithelial cells. Each cell has about 300 cilia, measuring between 5 and 10 μm in length and 0.1–0.3 μm in diameter. The cilia beat in regular waves, with a frequency of 10 Hz. Their role is to facilitate the movement of mucus from the nasal cavity to the nasopharynx, and ultimately to the GI tract, the combined effect being known as mucociliary clearance (Marttin et al 1998, Schipper et al 1991).

Mucociliary clearance is a non-specific defensive function which also presents a barrier to drug absorption. The mucous layer is normally 5–20 μm thick, consisting mostly of water containing glycoproteins, ions and various other proteins, such as enzymes and immunoglobulins. Glycoproteins give the mucus its viscous character, which causes foreign particles to become trapped, cleared to the GI tract and ultimately eliminated from the body. The mucus is actually divided into two layers, the one closest to the cell surface being a less viscous, watery substance. This aids clearance by lubricating the passage of the mucus over the cell surfaces and easing the action of cilia. The cilia work in a ratchet-like way by engaging the viscous outer gel layer of the mucus, moving it towards the nasopharynx, and then disengaging and returning to their starting position through the serous inner layer. The turnover time for mucus is variously quoted as approximately 10–15 minutes in total (Lee and Baldwin 1992) and 20 minutes for the half-life of mucociliary clearance (Schipper et al 1991).

Mucus presents a diffusional barrier to drug absorption, and any formulation must be able to overcome this, as well as remain in the region long enough to allow drug release and absorption. Various strategies for achieving these goals are given later, but it is important that any interruption to mucociliary clearance, whether from drug or excipient, should be minimal and temporary.

Nasal metabolism

The nasal route of administration avoids hepatic first-pass metabolism, but nasal mucosa does possess enzymatic activity as a protective mechanism against exogenous chemicals. Although not widely investigated yet, there is sufficient evidence to suggest that nasal first-pass metabolism may be a significant factor in the absorption of some drugs (Sarkar 1992). For example, there is a high content of cytochrome P450 enzymes; P450 monooxygenases can oxidize many nasally administered drugs, such as nasal decongestants and anaesthetics. Progesterone and testosterone have been found to be metabolized extensively in vitro, although in vivo they have nasal bioavailabilities approaching 100% (relative to i.v. administration). Uneven distribution of enzymes is thought to be the reason for this apparent anomaly, greater activity being found in the olfactory mucosa used in the in vitro studies, whereas in vivo the steroids were probably absorbed rapidly by the less active respiratory mucosa.

There are many other types of enzyme in the nasal mucosa which can act on conventional drugs. Examples include dehydrogenases, hydroxylases, carboxylesterases, carbonic anhydrase and various

phase II conjugative enzymes. Although these enzymes have usually been investigated for toxicological reasons, they may interfere with the efficient absorption of drugs. This is especially true for peptide- and protein-based drugs, as various enzymes, such as aminopeptidases, can significantly decrease the drug's bioavailability. It is now recognized that proteins and peptides are faced with substantial enzymatic, as well as physical, barriers to absorption across mucosal surfaces (Lee et al 1992).

The development of new nasal dosage forms should therefore include some consideration of the nature, extent and location of the drug's metabolism in the nose. Not all metabolism is undesirable, however, and certain enzymes, such as esterases, open the possibility of using prodrugs as a means of improving nasal delivery.

PHYSICOCHEMICAL FACTORS INFLUENCING DRUG ABSORPTION IN THE NASAL CAVITY

Investigation of how the physicochemical properties of various drugs influence their nasal absorption has provided some insight into the various mechanisms and routes of drug absorption. The most important properties are probably the size of the drug molecule, its charge and its degree of hydrophilicity (or lipophilicity).

Molecular size and weight

McMartin et al (1987) compiled literature data for over two dozen compounds ranging in molecular weight from 160 to 34 000 Da. Wherever possible, comparisons were made between absorption of the same compound after delivery by the nasal route in humans and rats and oral administration in humans. All three models showed similar effects: the absorption of small compounds (approx. 100 Da) was high, at around 80%, but this decreased markedly as molecular weight increased. Nasal absorption of a drug with a given molecular weight was slightly higher in rats than in humans, but the greatest difference was seen between oral absorption and nasal absorption in humans. Drug absorption after nasal administration was very much higher than after oral delivery of a particular compound. Also, the molecular weight cut-off, beyond which absorption was negligible, was about two orders of magnitude better for nasal administration (about 20 000 Da compared to around 200 Da for peroral delivery).

These data, even though they relate to a small group of compounds, show the utility of the nasal route, and they led to the proposal that the drugs (which tended to be hydrophilic) were absorbed by non-specific diffusion through aqueous channels between the cells, although other routes were a possibility. The molecular weight cut-off occurs because only molecules that are smaller than the channels can diffuse through them. The hypothesis that transport can occur through aqueous channels has received support in other studies using homologous series of hydrophilic compounds, such as poly(ethylene glycols) (Donovan et al 1990) and labelled dextrans with molecular weights between 1260 and 45 500 Da (Fisher et al 1992).

The effect of pH and the partition coefficient

Although molecular size is undoubtedly an important factor in influencing nasal absorption, the evidence in the preceding section tends to be based upon water-soluble compounds. More lipophilic compounds are likely to travel through one of the alternative routes, probably by partitioning across the mucosal cell membranes and diffusing through the cells at a rate slower than through intercellular channels. Most drugs can be ionized, and their partition coefficients are dependent upon environmental pH (unionized compounds will have a higher oil/water partition coefficient than ionized ones).

The pH at the surface of the mucosal cells has been reported to be 7.39 (Hirai et al 1981a); the mucus layer is slightly acidic, at pH 5.5–6.5 (Chien 1992). It should be mentioned that the local pH can be modified by a nasal formulation. Hirai et al (1981a) studied absorption in a perfused rat model of two model drug compounds at various pHs, a base (aminopyrine) and an acid (salicylic acid). The rate of absorption for both drugs increased as they became less ionized. Aminopyrine is absorbed more quickly at higher pHs, and the curve of absorption rate versus pH closely followed the degree of ionization versus pH curve. This suggests that the more lipophilic, unionized form of the drug is absorbed by passive diffusion through the mucosal cells. The salicylic acid was also absorbed more quickly in its unionized form (at low pH); the rate of absorption did not correspond with the degree of ionization: instead, it was higher than predicted. In this case it was suggested that there was another mechanism working alongside the slower diffusion through the cells, possibly the salicylic acid enhancing its own absorption. It is worth noting that despite both of the

model compounds being relatively small, there was no suggestion in this study of a fast absorption through intercellular channels.

Other mechanisms of drug absorption are also present in the nose. The absorption of benzoic acid decreases as pH increases, but there was still appreciable absorption at pHs where the acid is completely ionised (Huang et al 1985), suggesting the existence of two methods of absorption. It is possible that low pHs can have a direct action on the nasal mucosa: the absorption of secretin was greater at low pHs, and histological examination of the epithelial cells found changes in their structure at pH 3 (Ohwaki et al 1987).

The effect of pH on peptide drugs is more complex than on conventional drugs, as peptides have a large number of ionizable groups of either charge. Peptides are characterized by their isoelectric point, the pH at which they have no net charge and where their solubility is often lowest. The nasal absorption of insulin (isoelectric point at pH 5.4), as measured by the reduction in blood sugar in dogs, was greatest at a solution pH of 3.1 (net positive charge on insulin) and lowest at pH 6.1, near its isoelectric point. The absorption mechanisms of peptides are however, complex, owing to their size and structure, and it is unlikely that the effect of pH on insulin absorption is due to a more favourable partitioning of the protein. Indeed, there could well have been direct effects on the epithelial cells at pH 3.1.

The role of the partition coefficient so far has largely been inferred from the relative proportions of the ionized to the unionized forms of the drug, but this relationship is not pronounced. The absorption of a series of barbiturates was studied at pH 6.0, when they would be largely unionized, and only a fourfold variation in absorption was found despite a nearly 50-fold range in the chloroform/water partition coefficient (Huang et al 1985). An investigation into the dependence of the absorption of a series of progesterone derivatives on their octanol/water partition coefficients gave a similar finding: absorption increased with partition coefficient, but to a markedly lower extent (Corbo et al 1989). The latter workers determined partition coefficients in a nasal mucosa–buffer system and found a much better correlation with absorption than the conventional oil/water models give. It is clear from these findings that an appropriate measure of partitioning must be used in order to determine the role of partitioning in nasal absorption, but in general it appears that partitioning is rarely the only factor controlling absorption.

Other physicochemical factors affecting nasal absorption

The solubility of the drug and its dissolution rate are important, especially if it is presented as a solid dosage form (e.g. a powder), as it must be able to cross the mucous layer before it can be absorbed by the epithelial cells. In addition, powder morphology and particle size influence the deposition of the drug inside the nasal cavity. All of these factors must be considered in the design of formulations and will be addressed later in this chapter.

Physicochemical properties and mechanisms of absorption – a summary

The above discussion has shown that the influence of various properties on the absorption of drugs can provide valuable indicators about the mechanisms of drug absorption. It is also clear that the nasal absorption of drugs is a complicated affair, and there is rarely a single way in which a drug can be absorbed. Chien (1992) has written about the various absorption mechanisms; the summary given below is a broad generalization.

Aqueous channels between the cells provide a relatively good route for water-soluble compounds, with absorption being limited mostly by molecular size. Other drugs are absorbed by passive diffusion, possibly using a transcellular route, and there is evidence for active transport of some amino acids. Whatever the mechanism, combined literature data suggest that molecules with a molecular weight up to about 1000 Da should have a relatively good systemic bioavailability without the need for absorption promoters. Absorption promoters could increase this molecular weight limit to about 6000 Da, but beyond this weight it is unlikely that large molecules could be absorbed without causing unacceptable damage to the nasal cavity.

STRATEGIES FOR IMPROVING DRUG AVAILABILITY IN NASAL ADMINISTRATION

There are three main ways to maximize the systemic bioavailability of drugs administered nasally:

- Improve nasal residence time
- Enhance nasal absorption
- Modify drug structure to change physicochemical properties.

Increasing nasal residence time

Mucociliary clearance acts to remove foreign bodies and substances from the nasal mucosa as quickly as possible. One way of delaying clearance is to apply the drug to the anterior part of the nasal cavity, an effect which is largely determined by the type of dosage form used. The preparation could also be formulated with polymers such as methylcellulose, hydroxypropylmethylcellulose or polyacrylic acid (Carbopol), which increase the viscosity of the formulation and act as bioadhesives with the mucus (Schipper et al 1991). Increasing the residence time does not necessarily lead to increased absorption. This can be illustrated by considering insulin solutions with similar viscosities containing carboxymethylcellulose or Carbopol. The carboxymethylcellulose solutions do not enhance the absorption of insulin (Duchene et al 1988), whereas Carbopol solutions do (Morimoto et al 1985). Increasing the viscosity of solutions will decrease the rate of diffusion of molecules through them – hence the apparent lack of effect of carboxymethylcellulose – but the polymer may have other enhancing actions, such as opening the intercellular junctions, as has been suggested for Carbopol (Junginger 1990).

An interesting formulation development for lengthening residence time is the use of microspheres (Pereswetoff-Morath 1998). The absorption of insulin is increased by degradable starch microspheres which, although insoluble, swell in water and form a viscous, bioadhesive mass (Illum et al 1987). Subsequent studies indicate a more direct action of the microspheres on epithelial cells, whereby the swelling of the microspheres causes a temporary dehydration and shrinkage of the cells, followed by opening of the tight intercellular junctions (Edman et al 1992).

Enhancing nasal absorption

Absorption enhancers work by increasing the rate at which drugs pass through the nasal mucosa. Many act by altering the structure of the epithelial cells in some way, but they should accomplish this while causing no damage or permanent change. General requirements of an ideal absorption enhancer at its concentration in use are:

- It should give an effective increase in the absorption of the drug.
- It should not cause permanent damage or alteration to the tissues.
- It should not otherwise be irritant or toxic, either to the local tissues or to the rest of the body.
- It should be effective in small quantities.
- The enhancing effect should occur when absorption is required (i.e. there should not be a lag in its effect).
- The effect should be temporary and reversible.
- The enhancer should fulfil all other expectations of formulation excipients (e.g. stability and compatibility).

The major reason for developing and testing enhancers is to increase the absorption of peptides and proteins, because their size leads to a relatively poor bioavailability. There is a large body of work investigating enhancers for nasal delivery. Much of this is focused on peptide delivery, and the interested reader is directed to reviews such as those written by Behl et al (1998b), Chien (1992) and Hinchcliffe and Illum (1999) for more detail on the various classes of enhancer.

Surfactants and bile salts have received considerable attention. Hirai et al (1981b) investigated surfactants of many types, including non-ionic ethers and esters, and anionic surfactants, for their effect on insulin absorption and found them to be particularly effective enhancers. Unfortunately, the enhancement usually correlates with mucosal damage, as the surfactants associate with cellular components such as membrane lipids and proteins. In some cases the association is so severe as to cause extraction of lipids or proteins and loss of epithelial cells. Surfactants are therefore unsuitable for therapeutic use as enhancers, although Hinchcliffe and Illum (1999) suggest they are experimentally useful as reference compounds guaranteed to cause enhancement.

Bile salts have greater potential as they appear to possess much of the enhancing activity but less of the damage potential of surfactants (bile salts possess some surface activity and can form micelles). Commonly studied bile salts include sodium cholate, sodium deoxycholate, sodium glycocholate and glycodeoxycholate, sodium taurocholate and taurodeoxycholate, all of which can cause enhancement at concentrations of 10–20 mM (Behl et al 1998b).

Several mechanisms have been proposed for the enhancing action of bile salts:

- Increasing cell membrane permeability by forming temporary channels through the lipid structure;
- Forming intercellular aqueous pores by opening the tight junctions between cells;
- Increasing the lipophilicity of charged drugs by forming ion pairs;
- Inhibition of proteolytic enzymes.

The most likely of these mechanisms for increasing the absorption of peptides is the opening of intercellular channels, rather than increasing cell permeability; the latter would probably require massive disruption of the cell before substantial quantities of peptide could pass. Although bile salts are claimed to be safer than surfactants they can still cause damage to epithelial cells. Again there is a positive correlation between their enhancing activity and the damage caused.

Merkus et al (1993) have documented the damage potential of a number of bile salts and surfactants, using an index of morphological damage and a measurement of toxicity on the cilia (specifically the ciliary beat frequency, CBF). Decreases in CBF correspond to cellular damage, and measurement of CBF has the advantage of providing a relatively easy quantitative measure of toxicity which can be used to show the change in toxicity with time, including any reversal. Ciliotoxicity can also furnish another explanation for increased absorption, because a lower CBF tends to increase nasal residence time.

Sodium tauro-24,25-dihydrofusidate (STDHF) is an enhancer with a structure similar to bile salts that has been widely studied as an enhancer for proteins. It has good aqueous stability and solubility (>10% w/v) and is surface active, forming micelles at a critical micelle concentration of 2.5 mM. One study has found an approximately tenfold increase in the bioavailability of human insulin and increases in the bioavailability of labelled dextrans ranging from about fourfold (40 000 Da dextran) up to 72 times (4000 Da) (Lee and Baldwin 1992). The degree of penetration enhancement depends on the concentration of the STDHF, reaching a maximum at around 0.3%. As this is over the critical micelle concentration, any further increase in STDHF concentration will see a rise in the number of micelles, not individual enhancer monomers. It can be concluded that the monomer concentration is more important for producing the enhancing effect than the number of micelles. The same authors claim that STDHF produces only a temporary increase in absorption and causes relatively little cell damage, although Merkus et al (1993) show that STDHF is ciliotoxic at its optimum enhancing concentration. It should be noted, however, that the ciliotoxicity depends on the model used, with a much lower degree of toxicity in human tissue than in the animal model used (chicken trachea).

Another group of surface-active materials is the phosphatidylcholines, for example lysophosphatidylcholine. These are similar to compounds that occur naturally as part of cell membranes and, as might be expected, one mechanism of action of phosphatidylcholines is to disrupt the cell membrane and increase its permeability. They may also inhibit proteolytic enzymes, and lysophosphatidylcholine is mucolytic. They are effective enhancers of protein absorption and, despite their proposed mechanism of action, do not cause the expected damage to the nasal lining. Dodecanoyl-L-α-phosphatidylcholine (DPPC) is one derivative which has been developed with a high activity and low toxicity profile. It is undergoing development as an enhancer for insulin. Early studies showed that DPPC increases the absorption of insulin in humans with little or no irritation to the nose, although more recent work suggests that the amount of insulin absorbed by this route may not be therapeutically useful (Hinchcliffe and Illum 1999).

Cyclodextrins (CD) are hollow cylindrical molecules made up of glucose units in a cyclic arrangement: α-CD has six glucose units, β-CD has seven and γ-CD has eight. These compounds have found a wide range of pharmaceutical uses, from solubility enhancement to taste masking, because of their ability to form 'inclusion complexes'. Here, part or all of a drug molecule inserts itself in the hollow central cavity of the cyclodextrin molecule. The complexed drug molecule takes on some of the physicochemical properties of the cylcodextrin molecule. Derivatives of the cyclodextrins can be used to modify these properties. Cyclodextrins have polar outer surfaces and less polar interiors, so they tend to be water-soluble but have the ability to accommodate hydrophobic molecules as part of the inclusion complex. This results in an increased aqueous solubility for the included species.

Cyclodextrins increase the bioavailability of lipophilic compounds by increasing their aqueous solubility, and hence their availability, at the surface of the nasal epithelium (Merkus et al 1999). As an example, dimethyl-β-cyclodextrin (DMβCD) gives an absolute bioavailability for oestradiol of 95% in rats and 67% in rabbits, a three- to five-fold increase compared with a control preparation. This has been supported by clinical data in humans showing that nasally administered oestradiol is effective as an oestrogen-replacement therapy in ovariectomized postmenopausal women.

Data are available to show the enhancement of absorption of other lipophilic compounds and even hydrophilic drugs. Hydrophilic drugs which have an enhanced absorption include peptides (e.g. buserelin) and proteins (e.g. calcitonin). The bioavailability of insulin is not increased to a therapeutic level. Insulin is one of the major goals for therapeutic

administration via the nose, but this has yet to be achieved at sufficient levels even with the use of enhancers. The mechanism of enhancement of hydrophilic drug absorption by cyclodextrins has yet to be fully elucidated, but it is probably the result of a direct action on the nasal epithelium rather than a modification of the drug's physicochemical properties.

One of the main benefits of using cyclodextrins according to Merkus et al (1999) is their lack of toxicity, expressed as direct damage to the nasal cells, ciliotoxicity, or more systemic effects. Particularly useful are the methylated-β-cyclodextrins, which exhibit a combination of high activity and low toxicity.

The enhancers reported so far all act directly on the nasal epithelium, with a consequent risk of irritation and cellular damage. Alternative molecules, such as chitosan (a polysaccharide derived from the shells of crustacea) that have different mechanisms of delivery, have been studied (Illum 1998).

It is unlikely that there will be a single universal absorption enhancer and currently many of the most effective enhancers also cause damage to the nose. However, the volume of research and development work that has been conducted on these enhancers will make the nasal delivery of many drugs feasible.

Modifying drug structure

Structural modifications to the drug molecule are usually made in order to bestow more favourable physicochemical properties to the drug, for example increasing its aqueous solubility or improving its partitioning characteristics. Cyclodextrins can perform some of these functions, although they do not actually change the drug's structure. Some structural changes will be permanent, either by altering substituent groups on the molecule or by using different salt forms. However, they run the risk of changing the drug's therapeutic and toxicological profile and thus require regulatory approval, with its associated costs and lengthy studies.

An alternative is to use prodrugs, whereby the prodrug has favourable properties for absorption but is changed to the active drug on passing through the nasal epithelium. An example is the use of an ester (thereby increasing the drug's lipophilicity) that can be metabolized to the active drug by the esterases present in the nasal mucosa. Peptide and protein drugs especially may benefit from the formation of prodrugs (Krishnamoorthy and Mitra 1998).

NASAL DELIVERY SYSTEMS AND THEIR FORMULATION

General formulation issues

Nasal dosage forms must fulfil the functions of any other type of formulation. They must:

- be effective
- have an acceptable safety and stability, both chemical and microbiological
- be acceptable to the patient to ensure compliance.

If the formulation is a liquid it may commonly contain:

- antimicrobial preservatives (e.g. benzalkonium chloride)
- antioxidants (e.g. butylated hydroxytoluene)
- solubilizing agents or cosolvents (e.g. glycol derivatives)
- salts for adjusting pH and tonicity
- humectants, to minimize irritation to the nose (e.g. glycerol)
- viscosity-increasing agents (e.g. methylcellulose), and
- absorption enhancers.

Types of nasal dosage form and delivery system

The final dosage form used for nasal drug delivery is chosen after consideration of a wide range of issues, covering patient convenience, efficiency of drug delivery and formulation reasons. The specifics of the dosage form or delivery system plays a major role in the absorption of the drug by influencing its deposition. For example, as mentioned above, drugs deposited in the anterior part of the nasal cavity will be better absorbed than those applied further back.

Deposition mechanisms

There are three main ways of depositing inhaled particles on the nasal lining: impaction, sedimentation and diffusion (Kublik and Vidgren 1998).

Impaction Impaction occurs when there is a change in direction of the airflow – as happens when inspired air passes through the nasal valve – and the inertia of large or fast-moving particles carries them in their original direction. This is usually the main way of depositing particles in the turbulence caused by fast flow rates, or with particles larger than 0.5–1 μm. Varying the flow rate from the device or the aero-

dynamic particle size are the means by which the formulator can influence this type of deposition.

Sedimentation Sedimentation happens when the air is moving slowly and the particles settle slowly under the force of gravity. This mode of deposition is described by Stokes' equation. The only control the formulator can have over this practically is to ensure a slow flow rate and by modifying the drug particle size or formulation droplet size.

Diffusion The final method of deposition, diffusion, occurs by Brownian motion and is thus limited to very small particles (< 0.5 μm). Normally diffusion is not important, as the inspired particles are too large.

Nasal dosage forms

Nasal dosage forms will usually contain the drug in a liquid or powder formulation delivered by a pressurized or pump system. There have been several studies that have looked at the influence of dosage form and delivery system on the deposition and absorption of drugs, and useful summaries are given by Kublik and Vidgren (1998) and Su (1991).

Liquid formulation Liquid formulations are usually aqueous solutions of the drug and thus have the general benefits and drawbacks of pharmaceutical solutions. They are relatively simple to develop and manufacture compared to solid dosage forms, but often have a lower microbiological and chemical stability, requiring the use of various preservatives.

The liquid form can be soothing to the nasal lining but this may be countered by the excipients, such as the antimicrobial preservatives, which can cause irritation or ciliotoxicity. The simplest way to give a liquid is by nose drops, but their cheapness and convenience are usually outweighed by the inaccuracy of dosing volume and the likelihood of too-rapid clearance by the liquid running straight into the oesophagus. Dosing accuracy can be improved by using unit-dose packs containing a predetermined volume, but it is still the case that accurate placing of the drops requires some skill and dexterity on the part of the user.

Squeezed bottles Squeezed bottles are often used for nasal decongestants and work by spraying a partially atomized jet of liquid into the nasal cavity. They give a better absorption of drug by directing the formulation into the anterior part of the cavity and covering a large part of the nasal mucosa. Deposition and volume are still subject to the technique of the user – whether the patient squeezes the bottle smoothly or in a series of vigorous squirts, for example. This type of delivery system, where the formulation is expelled through a plain orifice without any type of valve system, is subject to contamination by 'suck-back', as external material can be drawn back into the container as the pressure on it is released.

Metered-dose pump systems Metered-dose pump systems offer greater control over dosing than any of the previous systems. They can deliver solutions, suspensions or emulsions, with a predetermined volume between 25 and 200 μL, thus offering deposition over a large area. The formulation scientist is able to incorporate a high degree of control over the size and localization of the dose by changing various factors, such as the rheological and surface tension characteristics of the formulation and the design and geometry of the pump.

The interactions between these factors are complex, but one example will show the type of control that is available. The angle at which the spray leaves the nozzle (the cone angle) will affect where the formulation is deposited. A small cone angle (about 35°) tends to be deposited towards the back of the nasal cavity and larger angles (tending towards 90°) will go further forward.

Metered-dose systems are also available as pressurized products. These permit the delivery of solid particulate preparations and present special formulation challenges because of the requirement to form a stable dispersion of the drug in the propellant system. Pressurized metered-dose systems give a good distribution of the formulation in the nasal cavity, but there is evidence that it is not as even as the distribution from metered-dose pumps.

Particle size and dose volume are two important factors for controlling delivery from metered-dose systems. The optimum particle size for deposition in the nasal cavity is 10 μm (mass median diameter; Su 1991). In addition, particulate formulations tend to have a longer residence time in the nasal cavity than liquids, because they are removed more slowly by mucociliary clearance.

The volume of formulation that can be delivered is obviously limited by the size of the nasal cavity, and larger volumes tend to be cleared faster despite covering a larger area. Better absorption is often achieved by administering two doses, one in each nostril, rather than a single large dose (e.g. two 80 μL doses rather than one 140 μL; Kublik and Vidgren 1998).

CONCLUDING REMARKS

The nasal route is one of many which is receiving a large amount of attention as a way of delivering

drugs into the systemic circulation. Attention is focused particularly on peptide and protein drugs, and several have reached the market. Delivery systems for these drugs will often need to incorporate absorption enhancers or some means of increasing the time that the drug is in contact with the nasal lining.

Recent advances uphold the promise of the nasal route as a convenient way to administer a wide range of drugs.

REFERENCES

Behl, C.R., Pimplaskar, H.K., Sileno, A.P., deMeireles, J., Romeo, V.D. (1998a) Effects of physicochemical properties and other factors on systemic nasal drug delivery. *Adv. Drug Del. Rev.*, **29**, 89–116.

Behl C.R., Pimplaskar H.K., Sileno A.P., Xia W.J., Gries W.J. deMeireles J., Romeo V.D. (1998b) Optimization of systemic nasal drug delivery with pharmaceutical excipients. *Adv. Drug Del. Rev.*, **29**, 117–133.

British National Formulary 39, March 2000. British Medical Association and Royal Pharmaceutical Society of Great Britain, London.

Chien, Y., (1992) *Novel Drug Delivery Systems*, 2nd edn, Marcel Dekker, New York, Ch 5 Nasal drug delivery.

Corbo, D.C., Huang, Y.C., Chien, Y.W. (1989) Nasal delivery of progestational steroids in ovariectomised rabbits 2. Effect of penetrant hydrophilicity. *Int. J. Pharm.*, **50**, 253–260.

Donovan, M., Flynn, G., Amidon, G. (1990) Absorption of polyethylene glycols 600 through 2000: the molecular weight dependence of gastrointestinal and nasal absorption. *Pharm. Res.* 7, 863–868.

Duchene, D., Touchard, F., Peppas, N.A. (1988) Pharmaceutical and medical aspects of bioadhesive systems for drug administration. *Drug Dev. Indus. Pharm.* **14**, 233–318.

Edman, P., Björk, E., Rydén, L. (1992) Microspheres as a nasal delivery system for peptide drugs. *J. Controlled Release*, **21**, 165–172.

Fisher, A., Illum, L., Davis, S., Schacht, E. (1992) Di-iodo-L-tyrosine labelled dextrans as molecular size markers of nasal absorption in the rat. *J. Pharm. Pharmacol.*, **44**, 550–554.

Hinchcliffe, M., Illum, L. (1999) Intranasal insulin delivery and therapy. *Adv. Drug Del. Rev.* **35**, 199–234.

Hirai, S., Yahika, T., Matsuzawa, T., Mima. H. (1981a) Absorption of drugs from the nasal mucosa of rat. *Int. J. Pharm.* 7, 317–325.

Hirai, S., Yashiki, T., Mima, H. (1981b) Mechanisms for the enhancement of the nasal absorption of insulin by surfactants. *Int. J. Pharm.*, **9**, 173–184.

Huang, C., Kimura, R., Nassar, R., Hussain, A.A. (1985) Mechanism of nasal absorption of drugs I: Physicochemical parameters influencing the rate of in-situ nasal absorption of drugs in rats. *J. Pharm. Sci.*, **74**, 608–611.

Illum, L. (1998) Chitosan and its use as a pharmaceutical excipient. *Pharm. Res.* **15**, 1326–1331.

Illum, L., Jorgensen, H., Bisgaard, H. (1987) Bioadhesive microspheres as a potential nasal drug delivery system. *Int. J. Pharm.*, **39**, 189–199.

Junginger, H.E. (1990) Bioadhesive polymer systems for peptide delivery. *Acta Pharm. Tech.*, **36**, 110–126.

Krishnamoorthy, R., Mitra, A.K. (1998) Prodrugs for nasal drug delivery. *Adv. Drug Del. Rev.*, **29**, 135–146.

Kublik, H., Vidgren, M.T. (1998) Nasal delivery systems and their effect on deposition and absorption. *Adv. Drug Del. Rev.*, **29**, 157–177.

Lee, W.A., Baldwin, P.A. (1992) Nasal membranes: structure and permeability. In: Junginger H.E. (ed) *Drug Targeting and Delivery: Concepts in Dosage Form Design*. Ellis Horwood, New York, Ch 16.

McMartin, C., Hutchinson, L.E.F., Hyde, R., Peters, G.E. (1987) Analysis of structural requirements for the absorption of drugs and macromolecules from the nasal cavity. *J. Pharm. Sci.* **76**, 535–540.

Marttin, E., Schipper, N.G.M., Verhoef, J.C., Merkus, F.W.H.M. (1998) Nasal mucociliary clearance as a factor in nasal drug delivery. *Adv. Drug Del. Rev.* **29**, 13–38.

Merkus, F.H.W.M., Schipper, N.G.M., Hermans, W.A.J.J., Romeijn, S.G., Verhoef, J.C. (1993) Absorption enhancers in nasal drug delivery: Efficacy and safety. *J. Controlled Release*, **24**, 201–208.

Merkus, F.H.W.M., Verhoef, J.C., Marttin, E., Romeijn, S.G., van der Kuy, P.H.M., Hermens, W.A.J.J., Schipper, N.G.M. (1999) Cyclodextrins in nasal drug delivery. *Adv. Drug Del. Rev.*, **36**, 41–57.

Morimoto, K., Morisaka, K., Kamanda, A. (1985) Enhancement of nasal absorption of insulin and calcitonin using polyacrylic acid gel. *J Pharm. Pharmacol.*, **37**, 134–136.

Mygind, N., Dahl, R. (1998) Anatomy, physiology and function of the nasal cavities in health and disease. *Adv. Drug Del. Rev.* **29**, 3–12.

Ohwaki, K., Ando, H., Kakimoto, F., Useugi, K., Watanabe, S., Miyake, S., Kayano, M. (1987) Effects of dose, pH, osmolarity on nasal absorption of secretin in rats: histological aspects of the nasal mucosa in relation to the absorption variation due to the effects of pH and osmolarity. *J. Pharm. Sci.*, **76**, 695–698.

Pereswetoff-Morath, L. (1998) Microspheres as nasal drug delivery systems. *Adv. Drug. Del. Rev.*, **29**, 185–194.

Sarkar, M.A. (1992) Drug metabolism in the nasal mucosa. *Pharm. Res.*, **9**, 1–9.

Schipper, N.G.M., Verhoef, J.C., Merkus, F.W.H.M. (1991) The nasal mucociliary clearance: relevance to nasal drug delivery. *Pharm. Res.*, 7, 807–814.

Su, K.S.E. (1991) Nasal delivery. In Lee, V.H.L. (ed) *Peptide and Protein Drug Delivery*. Marcel Dekker, New York, Ch 13.

33
Transdermal drug delivery

Brian Barry

CHAPTER CONTENTS

The main purpose of this chapter is to show how the physicochemical properties of a drug in a topical dosage form affect that drug's transdermal delivery (or percutaneous absorption). This process, and the drug's topical bioavailability, depend on the medicament leaving a formulation (cream, ointment, patch etc.) and penetrating through the stratum corneum into the viable epidermis and dermis. Within the living tissues the molecule usually produces its characteristic pharmacological response before the blood or lymph circulations remove it (possibly to produce a systemic effect). The ultimate aim in dermatological biopharmaceutics is to design active drugs, or prodrugs, and to incorporate them into vehicles or devices that deliver the medicament to the active site in the biophase at a controlled rate.

The plan of the chapter is:

- to introduce the reader to the structure, functions and topical treatment of human skin;
- to deal with the principles of membrane diffusion, skin transport, the properties influencing transdermal delivery, permeation with and without stratum corneum control, and the methods for studying these processes;
- to consider how to maximize the bioavailability of a topical drug;
- to review the philosophy of transdermal patches;
- to close with a brief discussion of dermatological vehicles and a protocol for producing a dermatological formulation.

The reader may then see that skin therapy is a paradox. At first sight it appears to be a simple form of treatment, yet closer examination reveals that sound dermatological design represents one of the most difficult aspects of the science of pharmaceutical formulation.

STRUCTURE, FUNCTION AND TOPICAL TREATMENT OF HUMAN SKIN

The skin combines with the mucosal linings of the urogenital, digestive and respiratory tracts to protect the internal body structure from a hostile external environment of varying pollution, temperature, humidity and radiation. The skin safeguards the internal organs, limits the passage of chemicals into and out of the body, stabilizes blood pressure and temperature, and mediates the sensations of heat, cold, touch and pain. It expresses emotions (such as the pallor of fear, the redness of embarrassment and

anger, and the sweating of anxiety). The integument identifies individuals through the characteristics particular to humans, e.g. colour, hair, odour and texture.

Skin damages easily: mechanically, chemically, biologically and by radiation. Thus the tissue suffers cuts, bruises, burns, bites and stings; detergents, chemical residues, organic solvents and pollutants attack and penetrate the surface; and micro-organisms and plants deliver contact allergens. Topical and systemic drugs, toiletries and cosmetics and many diseases may all harm the skin.

Anatomy and physiology

The human skin comprises three tissue layers: the stratified, avascular, cellular epidermis, the underlying dermis of connective tissue, and the subcutaneous fat (Fig. 33.1(a)). Hairy skin contains hair follicles and sebaceous glands; the glabrous skin of

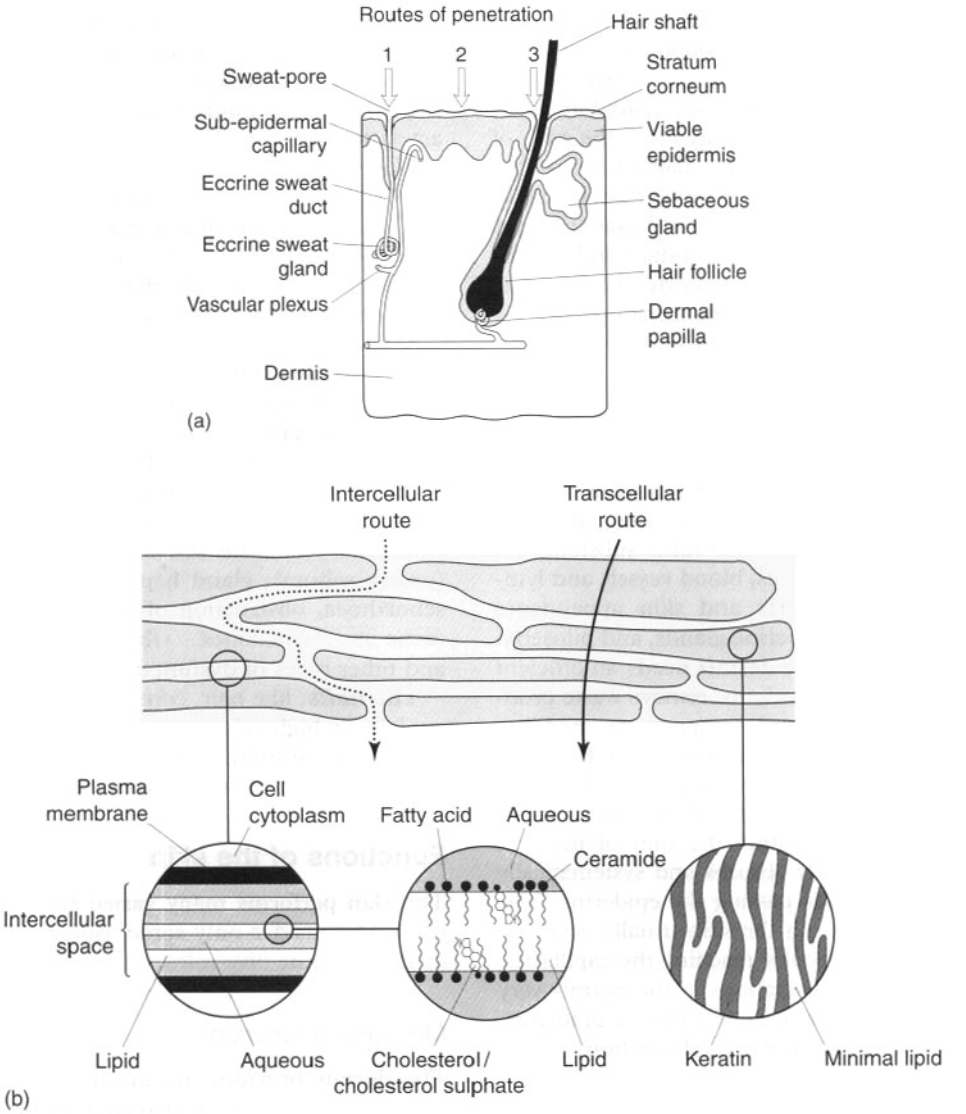


Fig. 33.1 Simplified diagram of skin structure and routes of drug penetration. (a) Macroroutes: (1) via the sweat ducts; (2) across the continuous stratum corneum; or (3) through the hair follicles with their associated sebaceous glands. (b) Representation of the stratum corneum membrane, illustrating two possible microroutes for permeation.

the soles and palms produces a thick epidermis with a compact stratum corneum, but there are no hair follicles or sebaceous glands.

The epidermis

The multilayered epidermis varies in thickness, ranging from about 0.8 mm on the palms and soles to 0.006 mm on the eyelids. The cells of the basal layer (stratum germinativum) divide and migrate upwards to produce the stratum corneum or horny layer. Humans survive in a non-aqueous environment because of the almost impermeable nature of this dead, dense layer, which is crucially important in controlling the percutaneous absorption of drugs and other chemicals. The stratum corneum may be only 10 μm thick when dry but swells severalfold in water. There are two main types of horny layer: the pads of the palms and soles, which are adapted for weight bearing and friction, and the remaining flexible, rather impermeable membranous layer. The basal cell layer also includes melanocytes, which produce and distribute melanin granules to the keratinocytes. Langerhans' cells (important in defence mechanisms operated by the immune system) are prominent in the epidermis.

The dermis

The dermis (or ***corium***), at 3–5 mm thick, consists of a matrix of connective tissue woven from fibrous proteins (collagen, elastin and reticulin) that are embedded in an amorphous ground substance of mucopolysaccharide. Nerves, blood vessels and lymphatics traverse the matrix and skin appendages (eccrine sweat glands, apocrine glands, and pilosebaceous units) pierce it. The dermis needs an efficient blood supply to convey nutrients, remove waste products, regulate pressure and temperature, mobilize defence forces and contribute to skin colour. Branches from the arterial plexus deliver blood to sweat glands, hair follicles, subcutaneous fat and the dermis itself. This supply reaches to within 0.2 mm of the skin surface, so that it quickly absorbs and systematically dilutes most compounds passing the epidermis. The generous blood volume in the skin usually acts as a 'sink' for diffusing molecules reaching the capillaries, keeping penetrant concentrations in the dermis very low, maximizing epidermal concentration gradients, and thus promoting percutaneous absorption.

The subcutaneous tissue

The subcutaneous fat (***subcutis***, ***hypoderm***) provides a mechanical cushion and a thermal barrier; it synthesizes and stores readily available high-energy chemicals.

The skin appendages

The ***eccrine sweat glands*** (2–5 million) produce sweat (pH 4.0–6.8) and may also secrete drugs, proteins, antibodies and antigens. Their principal function is to aid heat control, but emotional stress can also provoke sweating (the clammy palm syndrome).

The ***apocrine sweat glands*** develop at the pilosebaceous follicle to provide the characteristic adult distribution in the armpit (axilla), the breast areola and the perianal region. The milky or oily secretion may be coloured and contains proteins, lipids, lipoproteins and saccharides. Surface bacteria metabolize this odourless liquid to produce the characteristic body smell.

Hair follicles develop over all skin except the red part of the lips, the palms and soles, and parts of the sex organs. One or more ***sebaceous glands***, and in some body regions an apocrine gland, open into the follicle above the muscle that attaches the follicle to the dermoepidermal junction.

Sebaceous glands are most numerous and largest on the face, the forehead, in the ear, on the midline of the back and on anogenital surfaces; the palms and soles usually lack them. These holocrine glands produce sebum from cell disintegration; its principal components are glycerides, free fatty acids, cholesterol, cholesterol esters, wax esters and squalene. Abnormal sebaceous activity may lead to seborrhoea (excess sebum), gland hyperplasia without clinical seborrhoea, obstruction of the pilosebaceous canal (acne and comedones – whiteheads or blackheads), and other types of dysfunction – the dyssebacias.

The ***nails***, like hair, consist of 'hard' keratin with a relatively high sulphur content, mainly as cysteine. Unlike the stratum corneum, the nail behaves as a hydrophilic matrix with respect to its permeability.

Functions of the skin

The skin performs many varied functions but here we need consider only some aspects of its containment and protective roles.

Mechanical function

The dermis provides the mechanical properties of skin, with the epidermis playing a minor part. Skin is elastic, but once it has taken up its initial slack it extends further only with difficulty. With age, the skin wrinkles and becomes more rigid. The thin

horny layer is quite strong and depends for its pliability on a correct balance of lipids, water-soluble hygroscopic substances (the natural moisturising factor – NMF) and particularly water. The tissue requires some 10–20% of moisture to act as a plasticizer and so maintain its suppleness.

Protective function

Microbiological barrier The stratum corneum provides a microbiological barrier and the sloughing of groups of corneocytes (squames), with their adhering microorganisms, aids the protective mechanism. However, microbes penetrate superficial cracks and damaged stratum corneum may allow access to the lower tissues, where infection may develop. The so-called acid mantle (produced by sebaceous and eccrine secretions, at pH 4.2–5.6) probably does not defend the skin against bacteria via its acidity, as was once thought. However, skin glands also secrete short-chain fatty acids that inhibit bacterial and fungal growth. Nitric oxide, produced from nitrates in sweat, may help to prevent infections from skin pathogens, just as acidified nitrite has an antimicrobial action in the oral and gastrointestinal tracts. Bacteria are unlikely to enter the tiny opening of the inner duct of the eccrine gland; the entrances to the apocrine gland and the hair follicle are much wider, and these appendages may become infected.

Chemical barrier An important function of human skin is to bar the entry of unwanted molecules from outside while controlling the loss of water, electrolytes and other endogenous constituents. The horny layer is very impermeable to most chemicals and usually contributes the rate-limiting step in transdermal absorption. The intact skin is a very effective barricade because the diffusional resistance of the horny layer is large and the permeable appendageal shunt route provides only a small fractional area (about 0.1%).

Radiation barrier For skin exposed to sunlight, ultraviolet light of 290–400 nm is the most damaging. Three main acute reactions follow irradiation: erythema, pigmentation and epidermal thickening. Ultraviolet light stimulates melanocytes to produce melanin, which partially protects the skin. In a severe photosensitive disease such as xeroderma pigmentosum, sunlight may induce changes even in patients whose intense racial pigmentation makes them less susceptible to sunburn. Chronic reactions to sunlight include skin 'ageing', premalignancy and malignancy. Sun-damaged skin may produce solar keratoses, progressing to a squamous cell carcinoma. Bowen's disease, malignant melanoma and basal cell carcinoma may evolve.

Heat barrier and temperature regulation The stratum corneum is so thin over most body areas that it does not effectively protect the underlying living tissues from extremes of cold and heat; it is not an efficient heat insulator. The skin, however, is the organ primarily responsible for maintaining the body at 37 °C. To conserve heat, the peripheral circulation shuts down to minimize surface heat loss; shivering generates energy when chilling is severe. To lose heat, blood vessels dilate, eccrine sweat glands pour out their dilute saline secretion, water evaporates, and removal of the heat of vaporization cools the body.

Electrical barrier In dry skin, resistance and impedance are much higher than in other biological tissues.

Mechanical shock An acute violent blow bruises and blisters the skin; friction may blister or thicken the epidermis, producing callosities and corns. Accidental minor trauma to patients on corticosteroids may severely damage their skin, when the collagen is thinned by drug overuse.

Rational approach to drug delivery to and via the skin

There are three main ways to attack the problem of formulating a successful topical dosage form:

1. We can manipulate the barrier function of the skin: for example, topical antibiotics and antibacterials help a damaged barrier to ward off infection; sunscreen agents and the horny layer protect the viable tissues from ultraviolet radiation; and emollient preparations restore pliability to a desiccated horny layer.
2. We can direct drugs to the viable skin tissues without using oral, systemic or other routes of therapy.
3. The third approach uses skin delivery for systemic treatment. For example, transdermal therapeutic systems provide systemic therapy for conditions such as motion sickness, angina and pain.

Dermatologists aim at five main target regions: skin surface, horny layer, viable epidermis and upper dermis, skin glands and systemic circulation (Figs. 33.1 and 33.2).

Surface treatment

We care for the skin surface mainly by using a simple camouflage or cosmetic application, by forming a

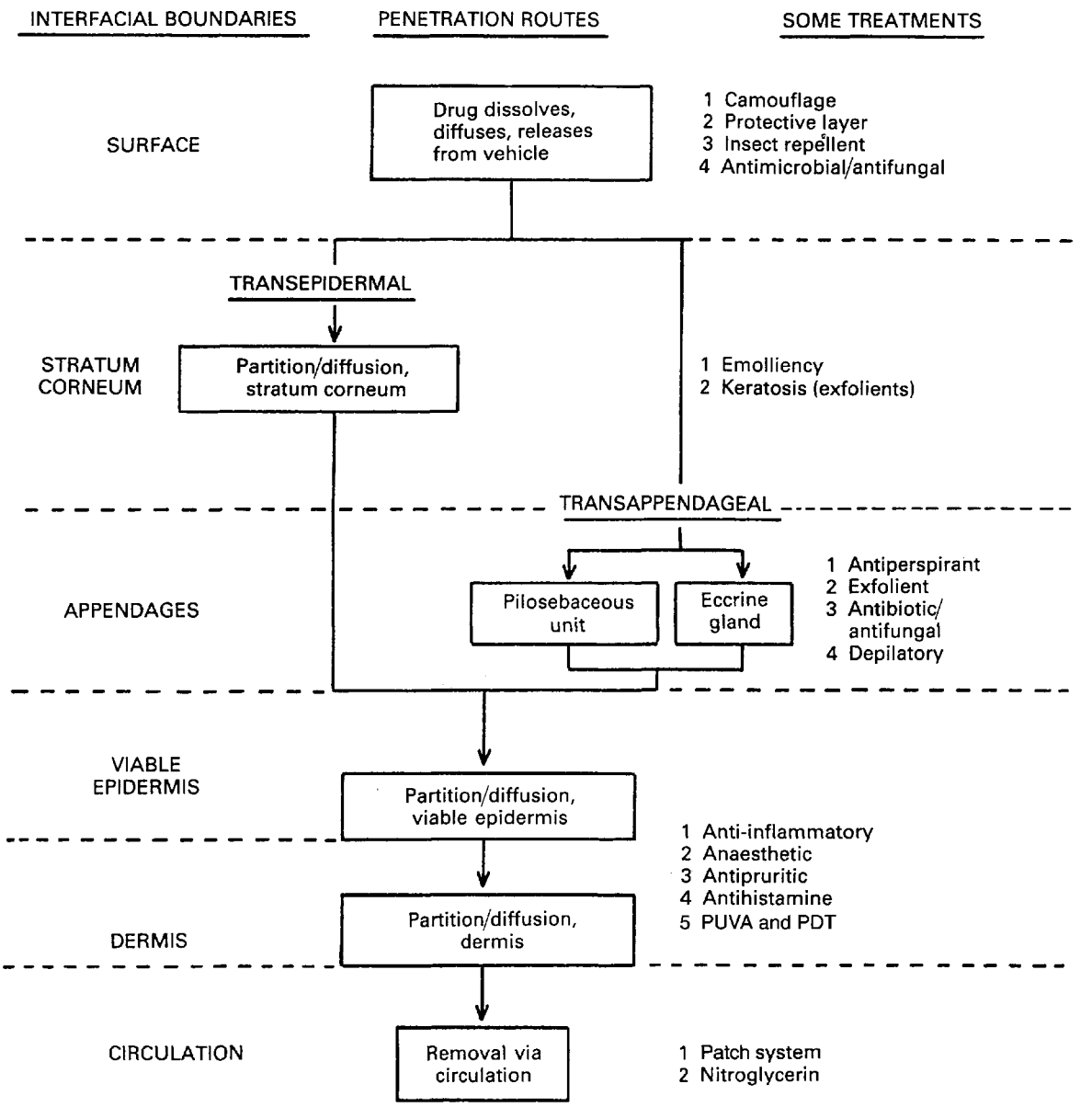


Fig. 33.2 The macroroutes by which drugs penetrate the skin and examples of treatments appropriate to disorders of the various strata (Reproduced with permission from Barry (1983).)

protective layer, or by attacking bacteria and fungi. Some examples include protective films, sunscreens, and barriers that hinder moisture loss and so avert chapping. For topical antibiotics, antiseptics and deodorants, the surface microorganisms are the target. Then, effective surface bioavailability requires that the formulation should release the antimicrobial so it can penetrate the surface skin fissures and reach the organisms. Developmental studies should at least confirm that the formulation releases and does not bind the medicament.

Stratum corneum treatment

The main therapies aimed at the horny layer improve emolliency by raising water content, or stimulate sloughing (keratosis) with, for example, salicylic acid. The insertion of moisturizing agents or kera-

tolytics into the stratum corneum involves their release from the vehicle and penetration into the tissue. Ideally, the medicament should not enter viable skin, but this is difficult to prevent.

Skin appendage treatment

We may reduce hyperhydrosis of the sweat glands with antiperspirants such as aluminium or other metal salts. In acne we use topical exfolients such as salicylic acid, tretinoin (retinoic acid) or isotretinoin, and benzoyl peroxide and antibiotics such as erythromycin and clindamycin. Other topical antibiotics applied in skin treatment include framycetin and neomycin sulphates, fusidic acid, polymyxins and mupirocin. Depilatories usually contain strontium or barium sulphides, or thioglycolates; male-pattern baldness may be treated with minoxidil and finasteride. Topical imidazoles (clotrimazole, econozole, miconazole, ketoconazole and sulconazole), amoralfine and terbinafine treat fungal diseases of the nails, stratum corneum and hair.

Delivering the medicament to the diseased site is a problem with appendage treatment. For example, it is difficult to achieve a high antibiotic concentration in a sebaceous gland when, as in acne, a horny plug blocks the follicle. When delivered through the skin, the drug may not be sufficiently hydrophobic to partition from the water-rich viable epidermis and dermis into the sebum-filled gland.

Viable epidermis and dermis treatment

We can treat many diseases provided that the preparation delivers drug to the receptor efficiently. However, many potentially valuable drugs cannot be used topically as they do not readily cross the stratum corneum. Hence, investigators may use stratagems such as adding penetration enhancers to diminish this layer's barrier function (see later). Another approach develops prodrugs, which reach the biological receptor and release the pharmacologically active fragment. The efficacy of many topical steroids depends partly on molecular groups that promote percutaneous absorption but which may not enhance drug-receptor binding.

Drug examples include topical steroidal and non-steroidal anti-inflammatory agents; corticosteroids may also be used in psoriasis. Antibiotics include those listed above. Anaesthetic drugs such as benzocaine, amethocaine and lignocaine reduce pain, and antipruritics and antihistamines alleviate itch, but they may cause sensitization. Topical 5-fluorouracil and methotrexate eradicate premalignant and some malignant skin tumours, and treat psoriasis. The psoralens (particularly in conjunction with ultraviolet light – PUVA therapy) mitigate psoriasis; and 5-aminolaevulinic acid (with visible light irradiation – photodynamic therapy) treats skin cancer.

Transcutaneous immunization

The skin has a highly effective immunological surveillance and effector system. A new therapy involves developing transcutaneous immunization via the topical application of vaccine antigens. The process uses an adjuvant such as cholera toxin added to a vaccine antigen (e.g. diphtheria toxoid) to induce antibodies to the diphtheria toxoid. The adjuvant and the antigen target Langerhans' cells, potent antigen-presenting cells in the epidermis. Simple application of the vaccine formulation to the skin of experimental animals and human volunteers has produced positive responses.

Systemic treatment via transdermal absorption

Generally, in the past healthy skin has not been used as a drug route during systemic attacks on disease, with the noteworthy exceptions of nitroglycerin and antileprotics. The body absorbs drugs slowly and incompletely through the stratum corneum, and much of the preparation is lost by washing, by adherence to clothes and by shedding with stratum corneum scales. Other problems include marked variations in skin permeability with regard to subject, site, age and condition, which make control difficult. However, in recent years considerable scientific work has led to the route being used to treat several conditions by means of transdermal patches (see later).

Figure 33.2 illustrates drug penetration routes and examples of treatments appropriate to various skin strata.

DRUG TRANSPORT THROUGH THE SKIN

Basic principles of diffusion through membranes

A useful way to study percutaneous absorption is first to consider how molecules penetrate inert (artificial) membranes, and then move on to the special situation of skin transport. An understanding of the basic principles of permeation through membranes is also valuable in all other areas of biopharmaceutics: oral,

buccal, rectal, nasal, lung, vaginal, uterine, injection or eye. The underlying mathematics are also relevant to dosage form design, particularly sustained- or controlled-release formulations and drug targeting.

The diffusion process

In passive diffusion matter moves from one region of a system to another, following random molecular motions. The basic hypothesis underlying the mathematical theory for isotropic materials (which have identical structural and diffusional properties in all directions) is that the rate of transfer of diffusing substance per unit area of a section is proportional to the concentration gradient measured normal to the section. This is expressed as Fick's first law of diffusion:

$$J = -D\frac{\partial C}{\partial x} \qquad (33.1)$$

where J is the rate of transfer per unit area of surface (the flux), C is the concentration of diffusing substance, x is the space coordinate measured normal to the section, and D is the diffusion coefficient. The negative sign indicates that the flux is in the direction of decreasing concentration, i.e. down the concentration gradient. In many situations D is constant, but in more complex materials D depends markedly on concentration; its dimensions are $(\text{length})^2 (\text{time})^{-1}$, often specified as $\text{cm}^2\ \text{s}^{-1}$.

Fick's first law contains three variables, J, C and x, of which J is additionally a multiple variable, dm/dt, where m is amount and t is time. We therefore usually employ Fick's second law, which reduces the number of variables by one. For the common experimental situation in which diffusion is unidirectional, i.e. the concentration gradient is only along the x-axis, Eqn 33.2 expresses Fick's second law as:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \qquad (33.2)$$

Many experimental designs employ a membrane separating two compartments, with a concentration gradient operating during a run and 'sink' conditions (essentially zero concentration) prevailing in the receptor compartment. If we measure the cumulative mass of diffusant, m, which passes per unit area through the membrane as a function of time, we obtain the plot shown in Figure 33.3. At long times the plot approaches a straight line and from its slope we obtain the steady flux dm/dt (Eqn 33.3):

$$\frac{dm}{dt} = \frac{DC_0K}{h} \qquad (33.3)$$

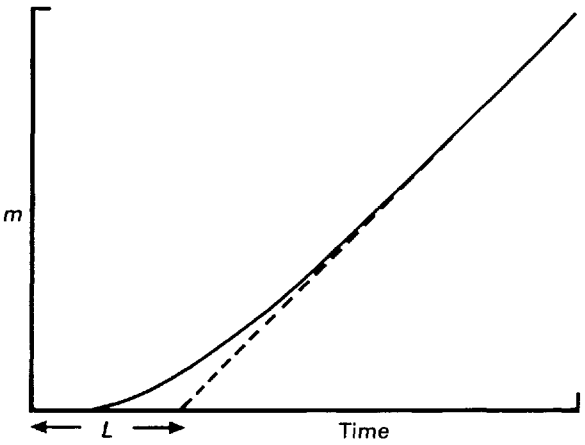


Fig. 33.3 The time course for absorption for the simple zero-order flux case obtained by plotting m, the cumulative amount of diffusant crossing unit area of membrane, as a function of time. Steady state is achieved when the plot becomes linear; extrapolation of the linear portion to the time axis yields the lag time L.

Here C_0 is the constant concentration of drug in the donor solution, K is the partition coefficient of the solute between the membrane and the bathing solution, and h is the thickness of the membrane.

If a steady-state plot is extrapolated to the time axis, the intercept so obtained at $m = 0$ is the lag time, L:

$$L = \frac{h^2}{6D} \qquad (33.4)$$

From Eqn 33.4, D is estimated provided that the membrane thickness, h, is available. Knowing these parameters and C_0, and measuring dm/dt, Eqn 33.3 provides one way of assessing K. Equation 33.3 shows why this permeation procedure may be referred to as a zero order process. By analogy with chemical kinetic operations, Eqn 33.3 represents a zero-order process with a rate constant of DK/h.

Sometimes with biological membranes (such as skin) we cannot separate the value of D from that of K. We then often use a composite parameter, the permeability coefficient, P, where $P = KD$ or $P = KD/h$. The latter definition is used when h is uncertain, e.g. diffusion through skin.

Complex diffusional barriers

Barriers in series The treatment above deals only with the simple situation in which diffusion occurs in a single isotropic medium. However, skin is a heterogeneous multilayer tissue and in percutaneous absorption the concentration gradient develops over

several strata. We can treat skin in terms of a laminate, each layer of which contributes a diffusional resistance, R, which is directly proportional to the layer thickness h and indirectly proportional to the product of the layer diffusivity D and the partition coefficient K with respect to the external phase. The total diffusional resistance of all the layers in a three-ply membrane such as skin (stratum corneum, viable epidermis and dermis) is given by the expression:

$$R_T = \frac{1}{P_T} = \frac{h_1}{D_1K_1} + \frac{h_2}{D_2K_2} + \frac{h_3}{D_3K_3} \qquad (33.5)$$

Here R_T is the total resistance to permeation, P_T is the thickness-weighted permeability coefficient, and the numerals refer to the separate skin layers.

If one segment has a much greater resistance than the other layers (e.g. the stratum corneum compared to the viable epidermis or dermis) then the single high-resistance phase determines the composite barrier properties. Then $P_T = K_1 D_1/h_1$, where the subscript 1 refers to the resistant phase.

Barriers in parallel Human skin is pierced by shunts and pores, such as hair follicles and sweat glands (Fig. 33.1). Investigators often idealize this complex structure and consider the simple situation in which the diffusional medium consists of two or more diffusional pathways linked in parallel. Then the total diffusional flux per unit area of composite, J_T, is the sum of the individual fluxes through the separate routes. Thus:

$$J_T = f_1 J_1 + f_2 J_2 \ldots \qquad (33.6)$$

where f_1, f_2 etc. denote the fractional areas for each diffusional route and J_1, J_2 etc. are the fluxes per unit area of each separate route. In general, for independent linear parallel pathways during steady-state diffusion

$$J_T = C_0 (f_1 P_1 + f_2 P_2 + \ldots) \qquad (33.7)$$

where P_1, P_2 ... represent the thickness-weighted permeability coefficients.

If only one route allows diffusant to pass, i.e. the other routes are impervious, then the solution reduces to the simple membrane model with the steady-state flux determined by the fractional area and the permeation rate through the open channel.

Skin transport

The skin is very effective as a selective penetration barrier. The epidermis provides the major control element – most small water-soluble non-electrolytes diffuse into the capillary system 1000 times more rapidly when the epidermis is absent, damaged or diseased. Furthermore, in the intact skin substances penetrate at rates that may differ 10 000-fold. It is important that pharmaceutical scientists predict and control this selective permeability by relating the physiological and physicochemical attributes of the skin to the properties of the penetrant in a vehicle. We need to correlate the intrinsic properties of the skin barrier with the molecular requirements for breaching it, as modified by interactions with the components of topical vehicles. The eventual aim in dermatological biopharmaceutics is to design drugs with selective penetrability for incorporation into vehicles or devices that deliver the medicament to the active site, at a controlled rate and concentration, for the necessary time.

Routes of penetration

When a molecule reaches intact skin it contacts cellular debris, microorganisms, sebum and other materials. The diffusant then has three potential entry routes to the viable tissue: through the hair follicles with their associated sebaceous glands, via the sweat ducts; or across the continuous stratum corneum between these appendages (see Fig. 33.1). We can summarize the relevant features before arriving at a general conclusion.

Sebum and surface material

The layer of sebum mixed with sweat, bacteria and dead cells is thin (0.4–10 μm), irregular and discontinuous; it hardly affects transdermal absorption. This lack of significant effect may seem surprising, particularly as over 300 volatile compounds have been detected on the skin surface, as well as many non-volatiles.

Skin appendages

Their fractional area available for absorption is small (about 0.1%) and this route usually does not contribute appreciably to the steady-state flux of a drug. However, the route may be important for ions and large polar molecules that cross intact stratum corneum with difficulty. Skin appendages may also act as shunts, important at short times prior to steady state diffusion, e.g. in bioassays that use pharmacological reactions. Thus minute concentrations of nicotinates or corticosteroids penetrating rapidly down the shunt route may quickly trigger erythema or blanching, respectively.

Although we usually ignore the hair follicle route for molecular flux under steady-state conditions,

very large molecules and particles of colloidal dimensions can target the follicle. Thus, 'naked' DNA has been used for immunization by topical application. It was speculated that normal follicles have efficient mechanisms for inducing immune responses to proteins in the follicle. A preparation made from antibodies from transgenic plants, when rubbed into the scalp, neutralized the hair-loss effects of toxic chemicals used in cancer chemotherapy. Colloidal particles, such as liposomes and small crystals, are useful for targeting the hair follicle. In general, particles larger than 10 μm remain on the skin surface, those between approximately 3 and 10 μm concentrate in the hair follicle, and when less than 3 μm they penetrate follicles and stratum corneum alike.

Epidermal route

The epidermal barrier function thus resides mainly in the stratum corneum. This has a 'bricks and mortar' structure, analogous to a wall (Fig. 33.1(b)). The corneocytes, consisting of hydrated keratin, comprise the 'bricks'. These are embedded in the 'mortar', which is composed of a complex lipid mixture of ceramides, fatty acids, cholesterol and cholesterol esters, formed into multiple bilayers. Most molecules penetrating through the skin use this intercellular microroute.

Because stratum corneum is dead, it is assumed that there are no active transport processes and no fundamental differences between in vivo and in vitro permeation processes. However, there may be discrepancies in how some substances permeate excised skin and skin in vivo. These differences may arise because we manipulate the skin to insert it into the diffusion apparatus, including possibly damaging it.

Topically applied agents such as steroids, hexachlorophane, griseofulvin, sodium fusidate and fusidic acid may form a depot or reservoir by binding within the stratum corneum.

Diseases that disrupt the horny layer, such as eczema and exfoliative dermatitis, may allow easier access.

The viable layers (particularly the epidermis) may metabolize and inactivate a drug, or activate a prodrug. The dermal papillary layer contains so many capillaries that the average residence time of a drug in the dermis may only be about a minute before it is washed away. Usually, the deeper dermal layers do not influence transdermal absorption, although drugs such as non-steroidal anti-inflammatory agents reach as far down as muscle. However, the dermis may bind a hormone such as testosterone, decreasing its systemic removal. If the penetrant is very lipophilic it crosses the horny layer to meet an aqueous phase, in which it is poorly soluble. The chemical potential immediately below the barrier may then become high, approaching that in the barrier. The potential gradient (stratum corneum to viable tissue) thus falls, together with the flux. The rate-determining step in percutaneous absorption then becomes barrier clearance, not barrier penetration.

Conclusions

The stratum corneum develops as a thin, tough, relatively impermeable membrane which usually provides the rate-limiting step in transdermal drug delivery. The entire horny layer, not just some specialized region, provides the diffusional resistance. The membrane allows no drug to pass readily, but nearly all low molecular weight molecules penetrate to some extent. The lipid bilayers of the intercellular route provide the main pathway. Diffusion is passive, governed by physicochemical laws in which active transport plays no part.

For electrolytes and large molecules with low diffusion coefficients, such as polar steroids and antibiotics, and for some colloidal particles, the appendages may provide the main entry route.

Once past the horny layer, molecules permeate rapidly through the living tissues and sweep into the systemic circulation.

The fraction of a drug that penetrates the skin via any particular route depends on:

- the physicochemical nature of the drug, particularly its size, solubility and partition coefficient;
- the timescale of observation;
- the site and condition of the skin;
- the formulation;
- how vehicle components temporarily change the properties of the stratum corneum.

PROPERTIES THAT INFLUENCE TRANSDERMAL DELIVERY

When a preparation is applied to diseased skin the clinical result arises from a sequence of processes:

1. Release of the medicament from the vehicle;
2. Penetration through the skin barriers;
3. Activation of the pharmacological response.

Effective therapy optimizes these steps as they are affected by three components, the drug, the vehicle and the skin.

Figure 33.4, which represents the movement of drug molecules arising from, for example, a transdermal drug delivery system with a rate-controlling membrane, illustrates the complexity of percutaneous absorption. Any drug particles must first dissolve so that molecules may diffuse towards the membrane within the patch. The penetrant partitions into the membrane, diffuses across the polymer and partitions into the skin adhesive. The molecules diffuse towards the vehicle/stratum corneum interface. They then partition into the stratum corneum and diffuse through it. Some drug may bind at a depot site; the remainder permeates further, meets a second interface, and partitions into the viable epidermis. For a lipophilic species this partition coefficient may be unfavourable, i.e. less than 1. Within the epidermis, enzymes may metabolize the drug or it may interact at a receptor site. After passing into the dermis, additional depot regions and metabolic sites may intervene as the drug moves to a capillary, partitions into its wall and out into the blood for systemic removal. A fraction of the diffusant may partition into the subcutaneous fat to form a further depot. A portion of the drug can reach deep muscle layers, as illustrated by, for example, the efficacy of non-steroidal anti-inflammatory drugs.

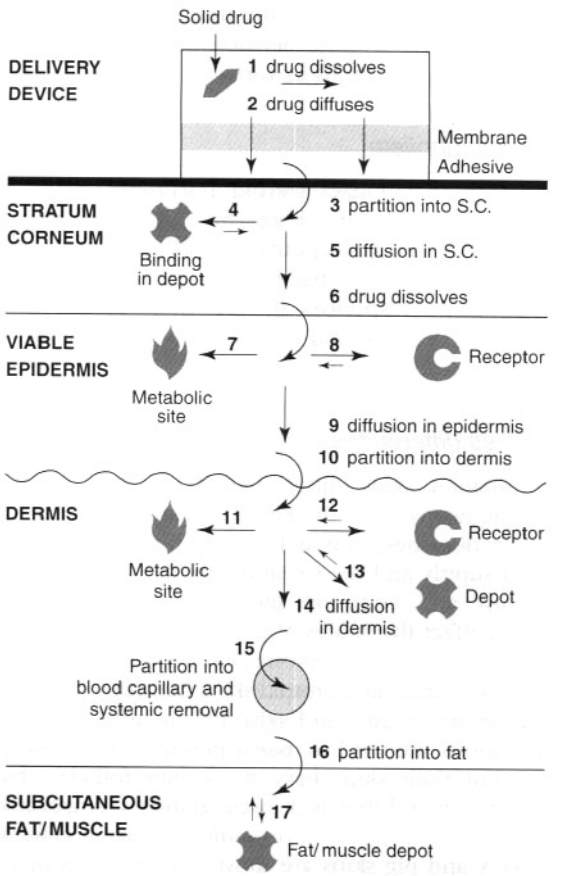


Fig. 33.4 Some stages in drug delivery from a transdermal patch. (Modified from Barry (1983), with permission.)

However, there are further complications. The following factors may be important: the non-homogeneity of the tissues; the presence of lymphatics; interstitial fluid; hair follicles and sweat glands; cell division; cell transport to and through the stratum corneum; and cell surface loss. The disease, the healing process, the drug and vehicle components may progressively modify the skin barrier. As vehicle ingredients diffuse into the skin, cellular debris, sweat, sebum and surface contaminants pass into the dermis, changing its physicochemical characteristics. Emulsions may invert or crack when rubbed in, and volatile solvents may evaporate.

To discuss this complicated process in a simple fashion, we review the material under the headings of biological factors and physicochemical factors. However, because transdermal delivery is a dynamic process it should be borne in mind that, as one variable changes, it usually causes several effects on drug flux. The various factors are separated for convenience, but in practice this is an artificial distinction, useful for discussion (and learning) purposes.

Biological factors

Skin condition

The intact, healthy skin is a tough barrier but many agents can damage it. Vesicants such as acids and alkalis injure barrier cells and thereby promote penetration, as do cuts, abrasions and dermatitis. In heavy industry, workers' skins may lose their reactivity or 'harden' because of frequent contact with irritant chemicals.

Many solvents open up the complex dense structure of the horny layer. Mixtures of non-polar and polar solvents, such as chloroform and methanol, remove the lipid fraction, forming artificial shunts through which molecules pass more easily.

Disease commonly alters skin condition; fortunately, for biopharmaceutical purposes we need only an elementary understanding of the gross changes in deranged skin. We are interested mainly in visible damage. Is the skin inflamed, with loss of stratum corneum and altered keratinization? Then

permeability increases. Is the organ thickened, with corns, calluses and warts, or as in ichthyosis? Drug permeation should now decrease. In diseases characterized by a defective stratum corneum, percutaneous absorption usually increases. Thus, a psoriatic plaque may take up twice as much 8-methoxypsoralen as does uninvolved skin.

After injury or removal of the stratum corneum, within 3 days the skin builds a temporary barrier that persists until the regenerating epidermis can form normal keratinizing cells. Even the first complete layer of new stratum corneum cells formed over a healing layer can markedly reduce permeation.

Skin age

It is often assumed that the skin of the young and the elderly is more permeable than adult tissue, but there is little evidence for any dramatic difference. Children are more susceptible to the toxic effects of drugs and chemicals, partly because of their greater surface area per unit body weight; thus potent topical steroids, boric acid and hexachlorophane have produced severe side-effects and death. Premature infants may be born with no stratum corneum. This can be turned to advantage by treating breathing difficulties with caffeine or pain with buprenorphine, via simple topical application instead of intravenous injection through tiny, delicate veins.

Blood flow

Theoretically, changes in the peripheral circulation could affect transdermal absorption; an increased blood flow could reduce the amount of time a penetrant remains in the dermis, and also raise the concentration gradient across the skin. Usually the effect is not clinically important (although it can be shown experimentally). In clinically hyperaemic skin, any increase in absorption almost always arises because the disease damages the skin barrier. Potent rubefacients, such as nicotinic acid esters, would also only have a significant effect after damaging the skin. Potent vasoconstricting agents, such as topical steroids, could reduce their own clearance rate or that of another drug.

Regional skin sites

Variations in cutaneous permeability around the body depend on the thickness and nature of the stratum corneum and the density of skin appendages. However, the absorption rate varies widely for a specific substance passing through identical skin sites in different healthy volunteers; the most permeable regions in some individuals compare with the least permeable sites in others. Investigators produce different rank orders for the permeabilities of skin sites in general; such permeabilities depend on both the intrinsic resistance to permeation per unit thickness of stratum corneum and the overall thickness of the tissue. As an example, plantar and palmar callus may be 400–600 μm thick compared to 10–20 μm for other sites. However, despite this greater thickness the tissue is less resistant ***per unit thickness***, so that the flux of diffusing drug is not so decreased compared to other sites, as might be expected.

Because of the relatively high permeability and ease of access of the site, the hyoscine Transderm system employs the postauricular skin (i.e. behind the ear) to insert drugs into the bloodstream. Originally this site was selected because it was thought that the layers of stratum corneum are thinner and less dense, there are more sweat and sebaceous glands per unit area, and many capillaries reach closer to the surface, increasing its temperature by 4–6°C relative to the thigh. However, facial skin in general is more permeable than other body sites, such as the limbs or torso.

Skin metabolism

The skin metabolizes steroid hormones, chemical carcinogens and some drugs. Such metabolism may determine the therapeutic efficacy of topically applied compounds (particularly prodrugs) and the carcinogenic responses in the skin. It has been estimated that the skin can metabolize some 5% of candidate topical drugs.

Species differences

Mammalian skins differ widely in characteristics such as horny layer thickness, sweat gland and hair follicle densities, and pelt condition. The capillary blood supply and the sweating ability differ between humans and common laboratory animals. Such factors affect the routes of penetration and the resistance to permeation. Subtle biochemical differences between human and animal skins may alter reactions between penetrants and skin. Frequently, mice, rats and rabbits are used to assess percutaneous absorption, but their skins have more hair follicles than human skin and they lack sweat glands. Comparative studies on skin penetration indicate that, in general, monkey and pig skins are most like that of humans. Hairless mouse skin has some similar characteristics: it has been widely used, but its stratum corneum is

fragile; hairless rat and fuzzy guinea pig may be better models for humans.

Snake skin has also been selected as it has the benefits that no sacrifice is required (the snake sheds its skin), it is readily available, and there is little leaching of chemicals to confuse UV assays.

Animal skin has been much used to obtain skin penetration data but it is best to use human skin whenever possible.

Physicochemical factors

Skin hydration

When water saturates the skin the tissue swells, softens and wrinkles and its permeability increases markedly. In fact, hydration of the stratum corneum is one of the most important factors in increasing the penetration rate of most substances that permeate skin. Hydration may result from water diffusing from underlying epidermal layers, or from perspiration that accumulates after the application of an occlusive vehicle or dressing. A dramatic example is the use of occlusive plastic films in topical steroid treatment, when the penetration of the steroid often increases tenfold. Occlusion decreases in the order: occlusive films = transdermal patches > lipophilic ointments > w/o cream > o/w cream. Powders, applied either as dusting powders or in lotions, provide a large surface area for evaporation and therefore dry the skin.

Commercial products are often promoted as skin softeners, with the presumption that they increase the skin's moisture content. However, they may contain humectants such as glycerol, propylene glycol or polyethylene glycol and emulsifiers that actually withdraw moisture from the skin.

Table 33.1 summarizes the effects that skin delivery systems may exert on stratum corneum hydration and permeability.

Temperature and pH

The penetration rate of material through human skin can change tenfold for a large temperature variation, as the diffusion coefficient decreases as the temperature falls. However, adequate clothing on most of the body would usually prevent wide fluctuations in temperature and penetration rates. Occlusive vehicles increase skin temperature by a few degrees, but any consequent increased permeability is small compared to the effect of hydration.

According to the simple form of the pH–partition hypothesis, only unionized molecules pass readily across lipid membranes. Now weak acids and bases dissociate to different degrees, depending on the pH and their pK_a or pK_b values. Thus, the proportion of unionized drug in the applied phase mainly determines the effective membrane gradient, and this fraction depends on pH. However, ionized molecules do penetrate the stratum corneum to a limited extent. Because they usually have a much greater

Table 33.1 Expected effects of skin delivery systems on horny layer hydration and skin permeability – in approximate order of decreasing hydration

Delivery system	Examples/constituents	Effect on skin hydration	Effect on skin permeability
Occlusive dressing	Plastic film, unperforated waterproof plaster	Prevents water loss; full hydration	Marked increase
Occlusive patch	Most transdermal patches	Prevents water loss; full hydration	Marked increase
Lipophilic material	Paraffins, oils, fats, waxes, fatty acids and alcohols, esters, silicones	Prevents water loss; may produce full hydration	Marked increase
Absorption base	Anhydrous lipid material plus water/oil emulsifiers	Prevents water loss; marked hydration	Marked increase
Emulsifying base	Anhydrous lipid material plus oil/water emulsifiers	Prevents water loss; marked hydration	Marked increase
Water/oil emulsion	Oily creams	Retards water loss; raised hydration	Increase
Oil/water emulsion	Aqueous creams	May donate water; slight hydration increase	Slight increase?
Humectant	Water-soluble bases, glycerol, glycols	May withdraw water, decreased hydration	Can decrease or act as penetration enhancer
Powder	Clays, organics, inorganics, 'shake' lotions	Aid water evaporation, decreased excess hydration	Little effect on stratum corneum

aqueous solubility than the neutral species, in saturated or near-saturated solutions, they may make a significant contribution to the total flux (i.e. although K may be small for ionized species, C_0 may be very high – see Eqn 33.3).

The stratum corneum is remarkably resistant to alterations in pH, tolerating a range of 3–9.

Diffusion coefficient

The diffusional speed of a molecule depends mainly on the state of matter of the medium. In gases and air, diffusion coefficients are large because the void space available to the molecules is great compared to their size, and the mean free path between molecular collisions is large. In liquids the free volume is much smaller, mean free paths are decreased and diffusion coefficients much reduced. In skin, the diffusivities drop progressively and reach their lowest values within the compacted stratum corneum matrix. For a constant temperature, the diffusion coefficient of a drug in a topical vehicle or in skin depends on the properties of the drug and the diffusion medium and on the interaction between them.

However, the ***measured*** value of D may reflect influences other than intrinsic mobility. For example, some drug may bind and become immobilized within the stratum corneum, and this process affects the magnitude of D as determined from the lag time (Eqn 33.4). However, regardless of such complications, the value of D measures the penetration rate of a molecule under specified conditions and is therefore useful to know.

Drug concentration

It was seen previously that the flux of solute is proportional to the concentration gradient across the entire barrier phase (Eqn 33.3). Thus, drug permeation usually follows Fick's law. One requirement for maximal flux in a thermodynamically stable situation is that the donor solution should be saturated. A formulator can optimize the solubility of a drug such as a corticosteroid by controlling the solvent composition of the vehicle. Then a saturated solution may be obtained at a selected concentration of the drug by experimenting with a series of solvents or, more usually, by blending two liquids to form a miscible binary mixture with suitable solvent properties.

Although the concentration differential is usually considered to be the driving force for diffusion, the chemical potential gradient or activity gradient is actually the fundamental parameter. Often the distinction is unimportant, but sometimes we must consider it. Thus, the thermodynamic activity of a penetrant in the donor phase or the membrane may be radically altered by, for example, pH change, complex formation, or the presence of surfactants, micelles or cosolvents. Such factors also modify the effective partition coefficient.

Partition coefficient

The partition coefficient (K, see Chapter 2) is important in establishing the flux of a drug through the stratum corneum (Eqn 33.3). When the membrane provides the sole or major source of diffusional resistance, then the magnitude of the partition coefficient is very important: this can differ by a factor of 10^8, drug to drug or (for one penetrant) vehicle to vehicle. The stratum corneum-to-vehicle partition coefficient is then crucially important in establishing a high initial concentration of diffusant in the first layer of the membrane.

Once it was incorrectly thought that good skin penetration required a K close to unity. However, many congeneric series of molecules display an ***optimal K***, well below which they are too water soluble to partition well into the horny layer. At higher values the compounds are so lipid soluble that they do not readily pass from the stratum corneum into the water-rich viable tissue. For a drug series this behaviour produces a parabolic or bilinear relation between pharmacological activity and partition coefficient.

Topical steroids provide a good example of the importance of the partition coefficient. Thus, triamcinolone is five times more active systemically than hydrocortisone, but it exhibits only about one-tenth the topical activity. Triamcinolone acetonide, with a more favourable K value, shows a 1000-fold increase in cutaneous activity. Betamethasone possesses only 10 times the topical potency of hydrocortisone, although it is some 30 times stronger systemically. Of the 23 esters of betamethasone tested, the 17-valerate has the highest topical activity and this coincides with the most balanced lipid/water partition coefficient. The anti-inflammatory responses to hydrocortisone and its C-21 esters behave similarly. As the side chain lengthens from 0 to 6 carbon atoms, the partition coefficient increases, as does the anti-inflammatory index. Thereafter the activity declines as the homologous series extends and the partition coefficient further increases.

Polar cosolvent mixtures, such as propylene glycol with water, may produce saturated drug solutions and so maximize the concentration gradient across the stratum corneum. However, the partition coefficient

of a drug between the membrane and the solvent mixture generally falls as the solubility in the solvent system rises. Thus, these two factors – increase in solubility and decrease in the magnitude of the partition coefficient – may oppose each other in promoting flux through the membrane, when the system is not saturated. Hence it is important not to oversolubilize a drug if the aim is to promote penetration: the formulation should be at or near saturation.

Surface activity and micellization affect transdermal delivery. There are two main complicating situations in membrane transport, one where the drug is surface active and forms micelles, and the other where additional surfactant is present. When the drug micellizes its total apparent solubility increases dramatically but the apparent partition coefficient decreases. However, the free monomer concentration remains constant, as does the true (monomer) partition coefficient. Then, if the micelle cannot cross the membrane this aggregate has little effect on the permeation process other than by serving as a reservoir to replace monomers as they leave to enter the skin, and by maintaining a constant donor concentration.

When drug and surfactant are present, the effect of surfactant on drug transport is complicated. The drug in the surfactant solution partitions between the micellar and the non-micellar pseudophases. The skin absorption of micellar drug may be negligible and the effective concentration of the unassociated diffusing species may be so lowered that its flux falls drastically. However, all these effects are more important for more permeable membranes such as gastrointestinal or buccal tissue.

Surfactants also have effects on skin that relate to the lowering of interfacial tension at the skin surface and in hair follicles, and changes in protein conformation and disruption of intercellular lipid packing in the stratum corneum. They then function as penetration enhancers – see later.

Complex formation is analogous in many ways to micellar solubilization in the manner in which it affects drug permeation. Thus, when complexes form, the apparent solubility and the apparent partition coefficient of the drug change. An increase in the apparent partition coefficient may promote drug absorption, e.g. some caffeine–drug complexes.

Molecular size and shape

Absorption is apparently inversely related to molecular weight: small molecules penetrate faster than large ones. However, the specific effect of the size of the penetrating molecule on the flux could only be determined if the effect of size could be separated from the resultant change in solubility characteristics. This is difficult to do, as the role of the partition coefficient is so dominant. With rare exceptions, we cannot experimentally increase the size of a molecule without also dramatically changing its partition coefficient.

It is even more difficult to determine the effect of molecular shape, separated from partition coefficient domination, and so nothing is known about this factor in skin permeation.

Ideal molecular properties for drug penetration

From the above considerations, we can deduce the ideal properties that a molecule would require so as to penetrate the stratum corneum well. In brief, these are:

- A low molecular mass, preferably less than 600 Da, when the diffusion coefficient will tend to be high;
- An adequate solubility in oil and water, so that the concentration gradient in the membrane can be high;
- A balanced partition coefficient;
- A low melting point; this correlates with good ideal solubility.

These features explain why transdermal patches can deliver adequate amounts of nicotine to be effective in smoking cessation therapy – this drug illustrates all these requirements well.

DRUG PERMEATION THROUGH SKIN

Stratum corneum rate controlling

It is useful here to summarize the basic concepts used in considering transdermal drug delivery, when the drug, the skin and the vehicle interact. Usually the relative impermeability of the stratum corneum provides the rate-limiting step in percutaneous absorption. The assumptions made in analysing relevant data include:

- Stratum corneum provides the rate-limiting step.
- Skin is a homogenous intact membrane; appendages are unimportant.
- Only a single non-ionic drug species is important, dissolving to form an ideal solution unaffected by pH, and dissolution is not rate limiting.

- Only drug diffuses from the vehicle. Formulation components neither diffuse nor evaporate, and skin secretions do not dilute the vehicle.
- Diffusion coefficient is constant with time or position in the vehicle or horny layer.
- Penetrant reaching viable tissue sweeps into the circulation, maintaining sink conditions below the stratum corneum.
- Donor phase depletes negligibly, i.e. constant drug concentration in the vehicle.
- Vehicle does not alter skin permeability during an experiment by, for example, changing stratum corneum hydration or by acting as a penetration enhancer.
- Drug remains intact and unaltered.
- Flux estimates are steady-state values.

Most analyses use Eqn 33.3, assume that h, the stratum corneum thickness, is constant, and relate changes in the penetration rate to variations in the other three parameters.

An important alternative form of Eqn 33.3 uses thermodynamic activities. Thus:

$$\frac{dm}{dt} = \frac{aD}{\gamma h} \qquad (33.8)$$

where a is the thermodynamic activity of the drug in its vehicle and γ is the effective activity coefficient in the skin barrier. To obtain the maximum rate of penetration, we see that we should use the drug at its highest thermodynamic activity. Now the dissolved molecules in a saturated solution are in equilibrium with the pure solid (which by definition is at maximum thermodynamic activity). The solute molecules are also thus at maximum activity. (It is useful in this context to think of thermodynamic activity as 'escaping tendency', i.e. the drive for the drug to escape from the vehicle and enter the skin). The conclusion, therefore, is that all vehicles that contain the drug as a finely ground suspension (in which the solute activity is maximal and equal to that of the solid) should produce the same penetration rate, ***provided that the assumptions above remain valid***. (The effect of supersaturation is dealt with later.)

If we use the placebo vehicle to dilute, for example, an ointment containing a drug in suspension, the drug flux (and hence the clinical efficacy) will not fall until the concentration decreases below the saturation value. This is why many extemporaneously diluted preparations have essentially the same therapeutic activity as their undiluted form – the dilution process never reached subsaturation.

Stratum corneum not rate controlling

Next we consider situations in which the impermeability of the stratum corneum is **not** important, i.e. we are concerned only with drug and vehicle interactions. Then the release of the drug from the vehicle provides the rate-limiting step and the skin functions as a sink. This could happen in a patient with a disrupted or absent horny layer, or when drug diffusion in the vehicle is exceptionally slow. The vehicle also provides the rate-controlling mechanism in many release studies that use either no membrane or an artificial porous membrane. Such experiments may correlate with clinical treatment only for patients with severely damaged skin.

When diffusion within the vehicle provides the rate-controlling step, our mathematical treatment assumes that the skin is a sink. Thus, it maintains essentially zero concentration of the penetrating material by passing it rapidly to the circulation. Then the concentration gradient develops solely in the applied formulation. Two important situations are absorption from solution and from suspension.

Absorption from solution: skin a perfect sink

We can deduce an equation that applies to the release of penetrant from one side of a layer of vehicle on the skin, under the following conditions:

1. Only a single drug species is important, it is in true solution, and it is initially uniformly distributed throughout the vehicle.
2. Only the drug diffuses out of the vehicle. Other components do not diffuse or evaporate, and skin secretions do not pass into the vehicle.
3. The diffusion coefficient does not alter with time or position within the vehicle.
4. When the penetrant reaches the skin, it absorbs instantaneously.

Under these limitations, Eqn 33.9 represents the relationship between m, the quantity of drug released to the sink per unit area of application, and C_0, the initial concentration of solute in the vehicle, D_v, the diffusion coefficient of the drug in the vehicle, and t, the time after application.

$$m \approx 2C_0\sqrt{\frac{D_v t}{\pi}} \qquad (33.9)$$

Differentiating this equation provides the release rate dm/dt:

$$\frac{dm}{dt} \approx C_0\sqrt{\frac{D_v}{\pi t}} \qquad (33.10)$$

Figure 33.5 illustrates plots of a typical release experiment for betamethasone 17-benzoate dissolved at various concentrations in a polar gel and diffusing into a chloroform sink. According to Eqn 33.9, a plot of m versus $t^{1/2}$ should provide a straight line, as Figure 33.6 illustrates. A relationship such as Eqn 33.9, or a modification in which m is still proportional to $t^{1/2}$, often fits data outside the limits used originally to define the equation, i.e. up to 65% release instead of only about 30%.

According to these equations we may alter the release rate of a drug from solution, and hence its bioavailability, by changing the drug concentration or the diffusion coefficient.

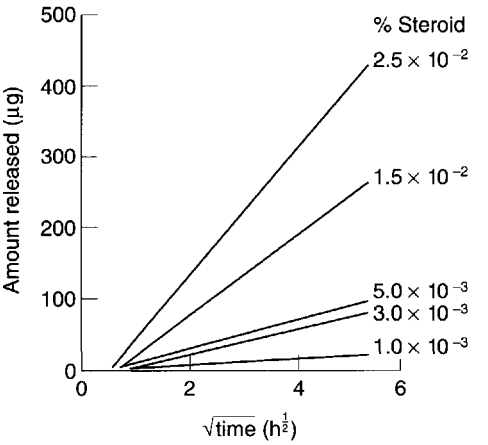


Fig. 33.6 In vitro release of betamethasone 17-benzoate from gel formulations as a function of the square root of time: steroid strength indicated on the plots. (Reproduced with permission from Barry (1983).)

Absorption from suspensions: skin a perfect sink

The amount and rate of release of a drug suspended in a vehicle, such as an ointment, may be related to time and to the variables of the system. The relevant equations are derived for a simple model system under the following conditions:

1. The suspended drug is micronized so that particle diameters are much smaller than the vehicle layer thickness.
2. The particles are uniformly distributed and do not sediment in the vehicle.
3. The total amount of drug, soluble and suspended, per unit volume (A) is much greater than C_s, the solubility of the drug in the vehicle.
4. The surface to which the vehicle is applied is immiscible with the vehicle, i.e. skin secretions do not enter the vehicle.
5. Only the drug diffuses out of the vehicle; vehicle components neither diffuse nor evaporate.
6. The receptor, which is the skin, operates as a perfect sink.

We can then obtain an equation that relates m to t in the form:

$$m = \sqrt{D_v t(2A - C_s)C_s} \qquad (33.11)$$

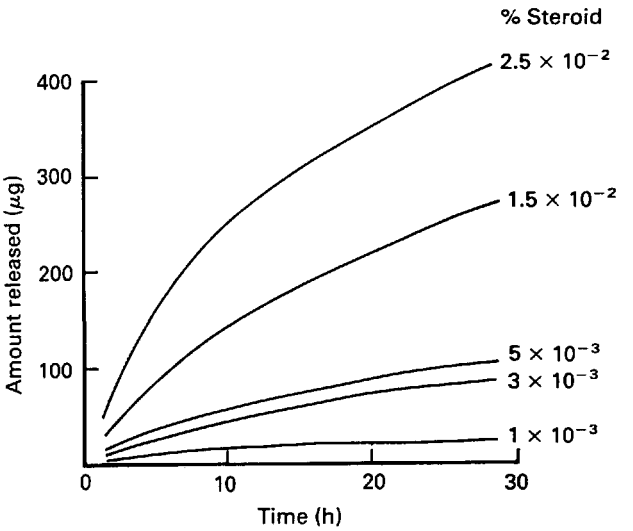


Fig. 33.5 In vitro release of betamethasone-17-benzoate from gel formulations as a function of time: steroid strength indicated on plots. (Reproduced from Barry (1983), with permission.)

This equation holds essentially for all times less than that corresponding to complete depletion of the suspended phase. If we differentiate Eqn 33.11 with respect to time, we obtain the instantaneous rate of release, dm/dt, given by:

$$\frac{dm}{dt} = \frac{1}{2}\sqrt{\frac{D_v(2A - C_s)C_s}{t}} \qquad (33.12)$$

For a common condition in which the solubility of the drug in the vehicle is very small and A is appreciable (i.e. $A >> C_s$), Eqn. 33.11 simplifies to:

$$m \approx \sqrt{2AD_vC_st} \qquad (33.13)$$

Then Eqn 33.12 becomes:

$$\frac{dm}{dt} \approx \sqrt{\frac{AD_vC_s}{2t}} \qquad (33.14)$$

These equations indicate that the formulator can manipulate drug bioavailability from ointment suspensions by altering the diffusion coefficient, the total concentration or the solubility. However, Eqn 33.14 predicts that $dm/dt \propto A^{1/2}$; doubling A only increases dm/dt by about 40%.

For obvious reasons Eqns 33.9–33.14 are often referred to as 'square root of time' relationships; they may also be called Higuchi equations, after the pharmaceutical scientist who developed them. The amount of drug released is proportional to the square route of time; the flux is an inverse function of time$^{1/2}$.

METHODS FOR STUDYING TRANSDERMAL DRUG DELIVERY

Experiments in percutaneous absorption may be designed to answer many questions, such as:

1. What is the drug flux through the skin and how do the apparent diffusion coefficient, partition coefficient, and structure–activity relationships control it?
2. What is the main penetration route – across the stratum corneum or via the appendages?
3. Which is more important clinically or toxicologically – transient diffusion (possibly down the appendages) or steady-state permeation (usually across the intact stratum corneum)?
4. Does the drug bind to the stratum corneum, the viable epidermis or the dermis; does it form a depot in the subcutaneous fat or penetrate to the deep muscle layers?
5. What is the rate-limiting step in permeation – drug dissolution or diffusion within the vehicle or patch; partitioning into, or diffusion through, the skin layers; or removal by the blood, lymph or tissue fluids?
6. How do skin condition, age, site, blood flow and metabolism affect topical bioavailability? Are differences between animal species important?
7. How do vehicles modify the release and absorption of the medicament? What is the optimal formulation for a specific drug – an aerosol spray, a solution, suspension, gel, powder, ointment, cream, paste, tape or delivery device?
8. Are vehicle components inert, or do they modify the permeability of the stratum corneum, if only by changing its hydration state?
9. To increase drug flux, should we use stratagems such as penetration enhancers, iontophoresis etc.?
10. Is the formulation designed correctly to treat intact stratum corneum, thickened epidermis or damaged skin?
11. Should the experimental design produce a pharmacokinetic profile, measuring absorption, distribution, metabolism and excretion in vivo?

No single method can answer all questions and provide a full picture of the complex process of transdermal absorption. We will therefore deal with the important general techniques, dividing them into in vivo and in vitro procedures. The former uses the skins of living humans or experimental animals in situ, whereas the latter employs isolated membranes and includes simple release studies.

In vitro methods

These are valuable techniques for screening and for measuring fluxes, partition coefficients and diffusion coefficients because the investigator can closely control laboratory conditions.

Excised skin

Excised skins from rats, mice and guinea pigs (normal and hairless), rabbits, hamsters, pigs, hairless dogs, snake, monkeys etc. have been mounted in diffusion cells. However, mammalian skin varies widely in stratum corneum properties and the number density of appendages. Thus, it is best to obtain human skin from autopsies, amputations or cosmetic

surgery. Investigators use either stratum corneum, epidermis, dermatomed skin or whole skin clamped in a diffusion cell. They measure the compound passing from the stratum corneum side through to a fluid bath. We can consider two main situations and illustrate just some of the many types of diffusion cell used.

In diffusion cells designed to examine steady-state flux and deduce fundamental parameters, a well-stirred donor solution at constant concentration releases penetrant through a membrane into an agitated 'sink' receptor liquid simulating the blood supply (Fig. 33.7). Figure 33.8 shows how three important quantities vary with time: the amount entering the membrane, that passing through (see also Fig. 33.3 and Eqns 33.3 and 33.4) and that remaining in the membrane.

Diffusion cells aimed more at simulating in vivo or clinical conditions use an agitated receptor solution to correspond to the blood and an unmixed donor phase to represent the formulation (Fig. 33.9). The donor compartment may be closed or open to ambient conditions or to controlled temperature and humidity; the skin may be washed, or materials added during an experiment. The test formulation may be a solid deposited from a volatile solvent, a liquid, a semisolid, a film or a drug device.

A technique known as attenuated total reflectance spectroscopy may also be used to measure passage across stratum corneum to determine the diffusion

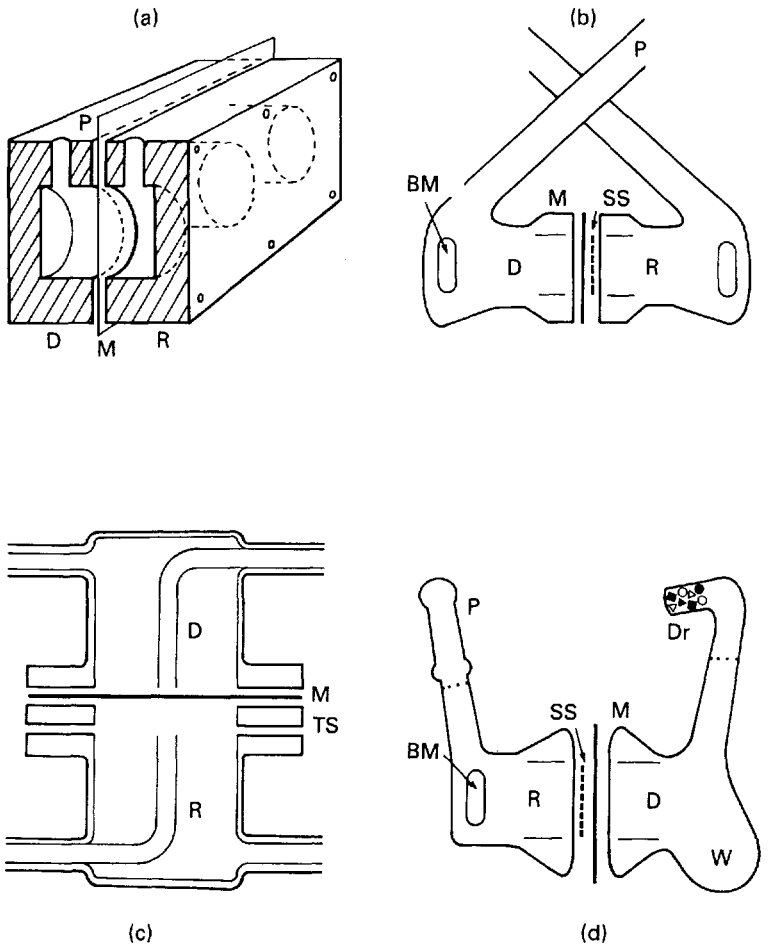


Fig. 33.7 Diffusion cells for zero-order or steady-state flux experiments (not to scale). (a) Bank of two cells drilled from a Perspex block. (b) Simple glass diffusion cell suitable for human skin. (c) Glass cell with continuously circulating donor and receptor solutions. (d) Glass cell used for determining vapour diffusion through the skin. D, donor compartment; R, receptor compartment; M, membrane; P, sampling port; BM, bar magnet; SS, stainless steel support; TS, Teflon support; W, well; Dr, drierite. (Reproduced with permission from Barry (1983).)

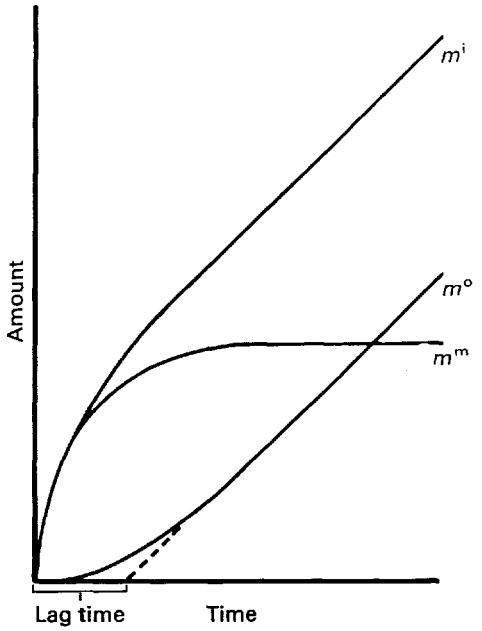


Fig. 33.8 Amount of penetrant entering the membrane (m^i), diffusing through (m^o) and being sorbed (m^m) under zero-order flux conditions. (Reproduced with permission from Barry (1983).)

coefficient of the penetrating molecule. The measuring technique may employ infrared or Raman spectroscopy.

Artificial membranes

Because human skin is variable and difficult to obtain, workers often use other materials, for example cellulose acetate, silicone rubber or isopropyl myristate; or lamellar systems designed to mimic the intercellular lipid of the stratum corneum. However, these membranes are not as complex as human skin and care must be taken if the results of such experiments are to be extrapolated to the clinical situation.

Release methods without a rate-limiting membrane

These procedures record drug release to a simple immiscible phase. They measure only those drug/vehicle interactions that affect release characteristics, and they do not determine skin absorption. Such procedures are mainly valuable in quality control protocols. Typical arrangements are shown in Figure 33.10; Eqns 33.9–33.14 may be used to analyse results.

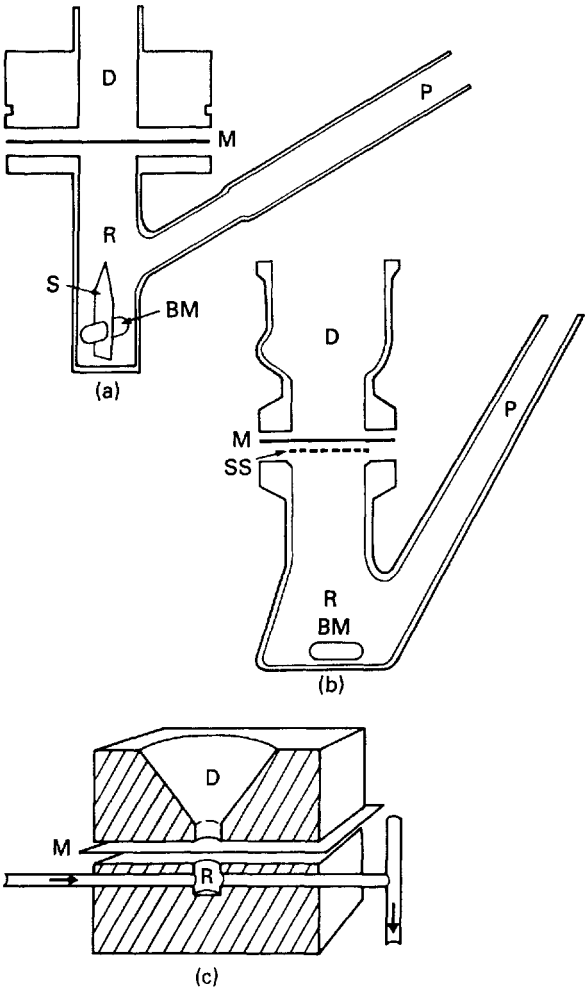


Fig. 33.9 Diffusion cells for simulation of in vivo conditions (not to scale). (a) Teflon and glass cell. (b) Glass cell with stainless steel support for the membrane. (c) Stainless steel cell with flow through receptor solution. D, donor compartment; R, receptor compartment; M, membrane; P, sampling port; BM, bar magnet; S, polyethylene sail; SS, stainless steel support. (Reproduced with permission from Barry (1983).)

In vivo methods

Often, in vivo methods use animals. However, most animals differ significantly from humans in features that affect percutaneous absorption: the thickness and nature of the stratum corneum, the density of hair follicles and sweat glands, the nature of the pelt, the papillary blood supply and biochemical aspects. Only a few techniques produce animal diseases that are similar to human afflictions. Thus animal models are valuable for studying the anatomy, physiology and biochemistry of skin, for screening topical

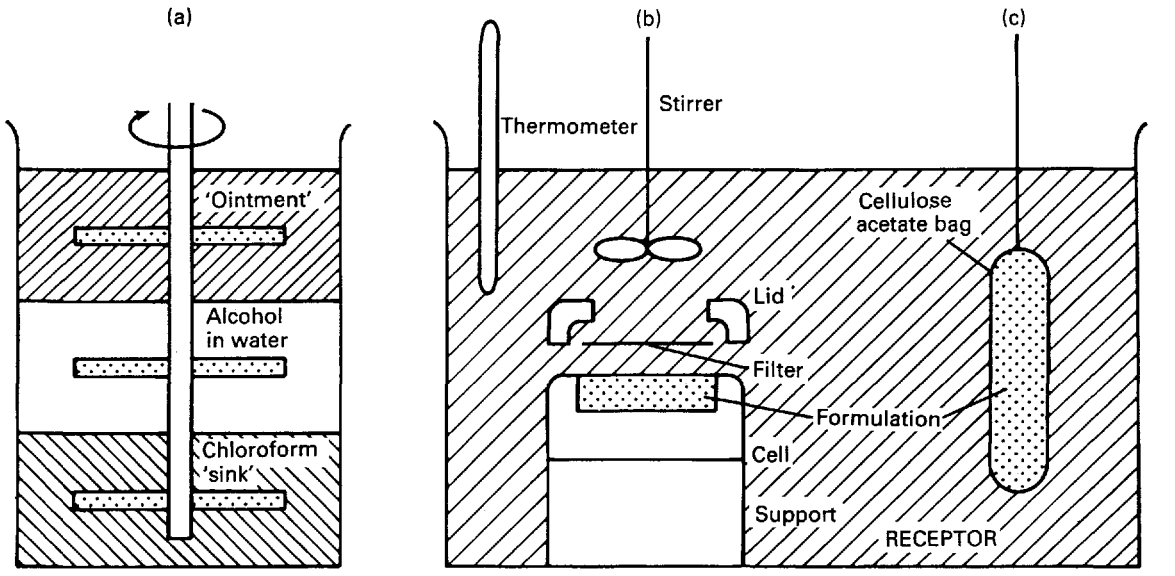


Fig. 33.10 Release methods without a rate-limiting membrane (not to scale). (a) Stirrer agitates three phases, which represent the formulation, the skin and the blood supply. (b) Release from an open container to a stirred immiscible receptor phase. (c) Release through a simple dialysis membrane. (Reproduced with permission from Barry (1983).)

agents, for detecting possible hazards, and for preliminary biopharmaceutical investigations. However, experience with animals cannot fully substitute for human studies; regulatory bodies will usually request additional data from human studies before granting a product licence.

Histology

Experimenters may try to locate skin penetration routes from microscopic sections; however, the cutting, handling and development of skin sections encourages leaching and the translocation of materials away from their original sites, a problem with histological methods in general.

Histochemical techniques have been used for those few compounds that produce coloured end-products after chemical reaction. A mistake in the past was to colour a penetrant with a dye and examine skin sections to locate the penetrant. However, each chemical species partitions and diffuses separately, and so the dyed complex dissociates and results are valid only for the dye itself. Added dye, or other different tracer molecules, should not be used.

A few compounds fluoresce, revealing their behaviour by microscopy, e.g. vitamin A, tetracycline and benzpyrene. Tumours fluoresce in photodynamic therapy when treated with 5-aminolaevulinic acid.

Microscopic autoradiography is difficult to apply to diffusable substances without modification. Substances emit α and β rays, and there may be considerable scattering on the autoradiogram; reducing substances reacting with the photographic emulsion – or an incorrect technique – can produce shadowing. Tritium-labelled isotopes are useful because of their weak emissions; strong β emitters darken areas up to 2 or 3 mm away, a great distance at the cellular scale.

Confocal microscopy can provide information at different depths in the skin.

Surface loss

In theory, we should be able to determine the flux of material into skin from the loss rate from the vehicle. However, because of skin impermeability, the concentration decrease in the vehicle would generally be small and analytical techniques would have to be sensitive and accurate. Also, those differences that could be detected would probably arise because the vehicle changed by evaporation or by dilution with sweat or transepidermal water, and not simply by drug partitioning into the skin. Alternatively, any drug decrease may only reflect deposition on the skin surface or combination with the stratum corneum, rather than penetration to the systemic circulation. Loss techniques have in the past been used mainly to monitor radioactive species; advances in HPLC analysis have made the methods more widely applicable.

Microdialysis

Microdialysis probes are inserted in the dermis and perfused with buffer. Drug molecules pass from the extracellular fluid into the buffer through pores in the membrane, which exclude large molecules, particularly proteins. The resulting drug solution is collected and analysed. For highly protein-bound drugs the technique requires very sensitive methods of analysis, as the drug concentrations in the samples are reduced accordingly.

Analysis of body tissues or fluids

When urinary analysis is used, the entire drug penetrating the skin should be accounted for by 'calibrating' the subject with a slow intravenous injection and a simultaneous determination of blood levels. This procedure allows for the pharmacokinetic factors inherent in drug absorption, distribution, storage, metabolism and excretion. Analysis of circulating blood can present difficulties with dilution, extraction and detection, although we may now routinely detect nanogram drug quantities. Faeces analysis alone has limited use. Sometimes the drug has an affinity for an animal organ, which can be removed and analysed, e.g. for iodine, iodides and mercury. Tissue biopsies may be analysed and even individual sections measured. Adhesive tape can strip sequential layers of the stratum corneum; the individual strips are then analysed for drug content.

Observation of a pharmacological or physiological response

If the drug stimulates a reaction in the viable tissues, we may use this to determine penetration kinetics. Local allergic, toxic or physiological reactions include sweat gland secretion, pigmentation, sebaceous gland activity, vasodilatation, vasoconstriction, vascular permeability, epidermal proliferation and keratinization. The most productive biopharmaceutical technique has been the vasoconstrictor or blanching response to topical steroids. For example, Figure 33.11 illustrates the blanching profiles of betamethasone benzoate in a quick-break aerosol foam, in the foam concentrate and in semisolids, compared with Betnovate Cream. The superiority of the aerosol foam and the inferiority of the benzoate cream are apparent. We use the vasoconstrictor test to screen novel synthetic steroids and develop topical formulations, to test marketed products for bioavailability and clinical efficacy, to perform fundamental studies in percutaneous absorption and to develop dosage regimens.

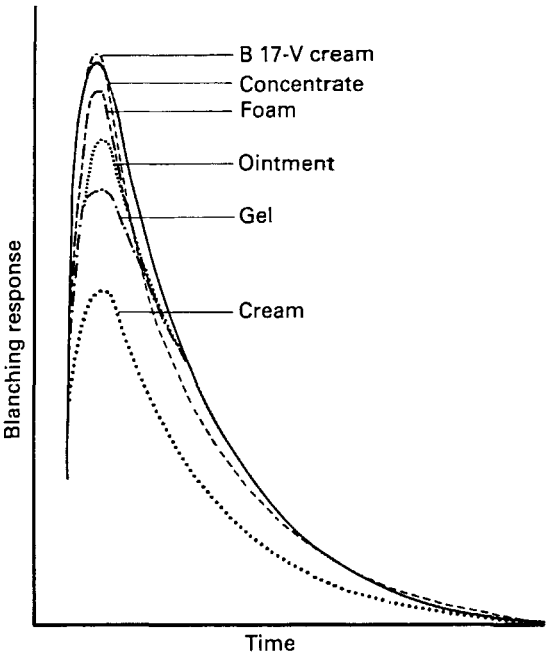


Fig. 33.11 Blanching response to betamethasone benzoate formulations containing 0.025% steroid and 0.1% betamethasone 17-valerate (B 17-V) cream. (Reproduced with permission from Barry (1983).)

Other response methods include changes in blood pressure (e.g. topical application of nitroglycerin), reduction in pain threshold, and the production of convulsions.

Physical properties of the skin

Relevant methods include the measurement of transepidermal water loss, thermal determinations (particularly differential scanning calorimetry), mechanical analysis, use of ultrasound, classification of function and dimension, spectral analysis (infrared and Raman), and the use of photoacoustic and electrical properties.

Bioassays

Many specialized bioassays screen topical formulations prior to clinical trial, including those for antibacterials, antifungals, antiyeast preparations, antimitotics, antiperspirants, sunscreen agents, antidandruff, anaesthetic–analgesic formulations, antipruritics, antiwart, poison oak/ivy dermatitis, antiacne and psoriasis. Topical corticosteroid bioassays are the most sophisticated and refined of all such bioassays (see previously in this chapter); other

steroid bioassays include antigranuloma, thymus involution, inflammation, cytological techniques and psoriasis bioassays.

MAXIMIZING THE BIOAVAILABILITY OF DRUGS APPLIED TO SKIN

Most drugs penetrate human skin poorly, and many major research efforts have attempted to maximize the input of such drugs. The great challenge for the future is to deliver the therapeutic peptides and proteins arising from the biotechnology revolution. The fundamental problem has two major aspects: not only is the stratum corneum a resistant barrier to penetration, but there is great biological variability in its impermeability. This section will therefore consider some ways in which pharmaceutical scientists have attempted to circumvent the horny layer barrier.

Drug or prodrug selection

The simplest way to consider factors that affect the permeation rate of a drug through the stratum corneum is via the simple equation for steady-state flux (Eqn 33.3). When, as is usual, the stratum corneum provides the major source of diffusional resistance of the skin, the partition coefficient of the drug is crucially important in establishing a high initial concentration in the first layer of this membrane. If the drug does not possess the correct physicochemical properties (usually it has too low a partition coefficient), we may be able to design a prodrug with an optimal partition coefficient for entering the skin barrier. After absorption and diffusion to the viable tissues, enzymes convert the prodrug to the active species. Very many steroid derivatives have topical anti-inflammatory activity greater than that of their parent steroids; with their optimized partition coefficients they act as prodrugs (see earlier in this chapter).

Chemical potential adjustment

One scheme for optimizing the bioavailability of a topical medicament is to ensure that the drug exhibits its maximum chemical potential, and thus its maximum thermodynamic activity, within the vehicle (Eqn 33.8).

Under ***ideal conditions***, the drug flux through the skin should be directly proportional to the drug activity in the vehicle, provided that D, γ and h remain constant. This has the important consequence that all saturated solutions of a specific drug in any vehicle (providing it does not modify the properties of the stratum corneum) should provide the same maximal flux.

However, it is possible to produce ***supersaturated*** solutions, either deliberately or via mixed vehicles evaporating on the skin. The theoretical maximum flux appropriate to thermodynamically stable vehicles may then increase manyfold using such supersaturated systems. Because supersaturated preparations are inherently unstable, manufacturers find it difficult to formulate them. An alternative is to employ controlled supersaturation in the rather special environments of transdermal patches; it is thought that this may offer the best hope of success. In practice, therefore, patients usually only meet the effects of supersaturation from volatile systems evaporating on their skin.

Hydration

Hydration of the stratum corneum is one of the most important factors in increasing the penetration rate of most substances: water opens up the compact layer of the horny layer. Moisturizing factors, occlusive films, hydrophobic ointments and transdermal patches all enhance skin bioavailability (see earlier, Skin hydration, and Table 33.1).

Ultrasound (phonophoresis)

This technique, used primarily in physiotherapy and sports medicine, involves placing the topical preparation on the skin over the area to be treated and massaging the site with an ultrasound source. The ultrasonic energy disturbs the lipid packing in the intercellular spaces of the stratum corneum by heating and cavitation effects, and thus enhances drug penetration into the tissue. A problem with the technique is, of course, the need for an ultrasonic probe, correctly focused to work in the stratum corneum. The method is therefore not readily suitable for home use.

Iontophoresis

Iontophoresis, the electrical driving of charged molecules into tissue, has applications in dentistry, ophthalmology, surgery and general medicine. As usually practised, the procedure involves passing a small direct current (approximately 0.5 mA/cm^2) through a drug-containing electrode in contact with the skin. A grounding electrode placed elsewhere on the body completes the electrical circuit. The trans-

port of the charged molecules is driven primarily by electrical repulsion from the driving electrode. However, polar neutral molecules can also be delivered by a current-induced convective flow of water (electro-osmosis). Considerable interest is now being shown in the possible transdermal delivery of therapeutic peptides and proteins, as well as many other drugs.

A problem with the technique is that, although the apparent current density per unit area is low, nearly all the current penetrates via the low-resistance route, i.e. the appendages, particularly the hair follicles. Thus the ***actual*** current density in the follicle may be high enough to damage growing hair. There is also concern about other possible irreversible changes to the skin.

As with ultrasound there is the problem of home use, although considerable work has been done on miniaturizing systems for patient use, e.g. paper batteries and wristwatch-like devices.

Electroporation

Electroporation is the creation of aqueous pores in the lipid bilayers by the application of short (micro- to millisecond) electrical pulses of approximately 100–1000 V/cm. Flux increases of up to 10 000-fold have been obtained for charged molecules. Again, there is the problem of instrumentation for home use for this potent technique.

Electroporation may combine with iontophoresis to enhance the permeation of peptides such as vasopressin, LHRH (luteinizing hormone-releasing hormone), neurotensin and calcitonin.

Stratum corneum removal

Laser ablation uses high-powered pulses from a laser to vaporize a section of the horny layer so as to produce permeable skin regions. The apparatus is costly and requires expert operation to avoid damage such as burns – it is hardly appropriate for home use. Hot needles have also been proposed.

Adhesive tape can remove the horny layer prior to drug application, as can controlled mechanical dermal abrasion. One other method forms a bleb (blister) by suction, an epidermatome removes the raised tissue, after which a morphine solution delivered directly to the exposed dermis produces fast pain relief.

Photomechanical wave

A drug solution is placed on the skin, covered by a black polystyrene target, and irradiated with a laser pulse. The resultant photomechanical wave produces stresses in the horny layer that enhance drug delivery. The technique is likely to remain experimental.

Needle array

The stratum corneum can be simply bypassed by an injection, and one development of this approach is a device consisting of 400 microneedles to insert drug just below the barrier. The feel to the skin is rather like a cat's tongue, or sharkskin. The solid silicon needles (coated with drug) or hollow metal needles (filled with drug solution) penetrate the horny layer without breaking and without stimulating nerves in the deeper tissues. Flux increases of up to 100 000-fold are claimed. The technique may also be combined with iontophoresis.

Penetration enhancers

Substances exist which temporarily diminish the impermeability of the skin. Such materials, known also as ***accelerants*** or ***sorption promoters***, if they are safe and non-toxic, can be used clinically to enhance the penetration rate of drugs and even to treat patients systemically by the dermal route. The attributes of the ideal penetration enhancer are:

1. The material should be pharmacologically inert.
2. It should be non-toxic, non-irritating and non-allergenic.
3. The action should be immediate and the effect should be suitable and predictable.
4. Upon removal of the material, the skin should immediately and fully recover its normal barrier property.
5. The enhancer should not cause loss of body fluids, electrolytes or other endogenous materials.
6. It should be compatible with all drugs and excipients.
7. The substance should be a good solvent for drugs.
8. The material should be cosmetically acceptable (good spreadability and skin 'feel').
9. The chemical should formulate into all the variety of preparations used topically.
10. It should be odourless, tasteless, colourless and inexpensive.

No single material possesses all these desirable properties. However, very many substances exhibit several of these attributes and they have been investigated clinically or in the laboratory. A sample summary includes:

- Water
- Sulphoxides (especially dimethylsulphoxide) and their analogues
- Pyrrolidones
- Fatty acids and alcohols
- Azone and its derivatives
- Surfactants – anionic, cationic and non-ionic
- Urea and its derivatives
- Alcohols and glycols
- Essential oils, terpenes and derivatives
- Synergistic mixtures.

For safety and effectiveness, the best penetration enhancer of all is water. Most substances penetrate better through hydrated stratum corneum than through the dry tissue. Thus, any chemical which is pharmacologically inactive, non-damaging, and which promotes horny layer hydration, can be considered as a penetration enhancer. Examples include the natural moisturizing factor and urea.

The literature on the use of skin penetration enhancers is voluminous – entire books have been devoted to considering various theories and examples. One simple way to classify enhancer actions is via the ***lipid–protein–partitioning*** concept. This hypothesis suggests that accelerants act in one or more ways selected from three main possibilities (see Fig. 33.1(b)).

Lipid action The enhancer interacts with the organized intercellular lipid structure of the stratum corneum so as to disrupt it and make it more permeable to drug molecules. Very many enhancers operate in this way. Some solvents may act by extracting the lipid components, thus making the horny layer more permeable.

Protein modification Ionic surface active molecules in particular tend to interact well with the keratin in the corneocytes, to open up the dense keratin structure and make it more permeable. However, the intracellular route is not usually important in drug permeation, although ***drastic*** reductions to this route's resistance could open up an alternative pathway for drug penetration.

Partitioning promotion Many solvents can enter the stratum corneum, change its solvent properties, and thus increase the partitioning of a second molecule into the horny layer. This molecule may be a drug, a coenhancer or a cosolvent. For example, ethanol has been used to increase the penetration of the drug molecules nitroglycerin and oestradiol. Propylene glycol is also widely used, particularly to provide synergistic mixtures with molecules such as Azone, oleic acid and the terpenes, i.e. to raise the concentration of these enhancers in the horny layer.

Ion pairs

Charged molecules do not readily penetrate the stratum corneum, so investigators have borrowed a technique used in analytical science. This is to form a lipophilic ion pair by adding a suitable species of a charge opposite to that of the drug ion. The complex formed then readily partitions into the lipid of the stratum corneum, as the charges temporarily neutralize each other. The ion pair diffuses across the horny layer to meet the aqueous viable epidermis. There the complex dissociates into its component charged species, which are readily soluble in water and thus partition into the epidermis and diffuse onward. However, the magnitude of the enhancement obtained is not great, typically being only about twofold.

Complex coacervates

Complex coacervation is the separation of oppositely charged ions into a dense coacervate oil phase, which is rich in the ionic complex. A coacervate may be thought of in this context as a further development of ion pairs. Similarly to ion pairs, the coacervate partitions into the stratum corneum, where it behaves as ion pairs, diffusing, dissociating and passing into the viable tissues. As for simple ion pairs, the flux enhancement obtained is similarly modest.

Liposomes and transfersomes

Liposomes are colloidal particles, typically consisting of phospholipids and cholesterol, to which other ingredients may be added. These lipid molecules form concentric bimolecular layers in the form of vesicles, that may be used to entrap and deliver drugs to and through the skin. How well they transport drugs ***through*** the skin is controversial, although significant enhanced fluxes, compared to saturated aqueous solution (maximum thermodynamic activity), have been obtained for several drugs. For example, the flux of oestradiol has been increased 20-fold using a variety of liposomes. The same vesicles increased 10-fold the deposition of this hormone within the stratum corneum.

Transfersomes are a special type of liposome that incorporate so-called 'edge activators' – molecules such as sodium cholate. The inventors claim that such vesicles are ultradeformable (up to 10^5 times that of an unmodified liposome). As such they can squeeze through pores in the stratum corneum which are less than one-tenth the diameter

of the liposome. Thus, sizes up to 200–300 nm can penetrate intact skin. Two particular features are claimed to be important when using transfersomes: they require a hydration gradient to encourage skin penetration, so that they only exhibit their marked delivery abilities when applied to the skin under non-occluded conditions. Then the hydration gradient operating from the (relatively) dry skin surface towards the waterlogged viable tissues drives the transfersomes through the horny layer (Fig. 33.12). Secondly, they work best under in vivo conditions.

Truly remarkable results are claimed for transfersomes. Data indicate that as much as 50% of a topical dose of a protein or peptide (such as insulin) penetrates the skin in vivo in 30 minutes.

High-velocity particles

The PowderJect system fires solid particles through the stratum corneum into the lower skin layers, using a supersonic shockwave of helium gas travelling at Mach 2–3. The claimed advantages of the system include:

- pain-free delivery – the particles are too small to trigger the pain receptors in the skin;
- improved efficacy and bioavailability;
- targeting to a specific tissue, such as a vaccine delivered to epidermal cells;
- sustained release, or
- fast release;
- accurate dosing;
- overcomes needle phobia;
- safety – the device avoids
  - skin damage
  - infection from needles or splashback of body fluids, particularly important for HIV and hepatitis B virus.

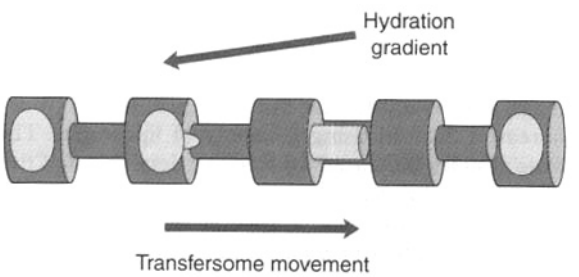


Fig. 33.12 Ultradeformable *transfersome* squeezing through minute pores in the stratum corneum, driven by the water concentration gradient. The modified liposome thus penetrates from the horny layer surface (relatively dry) to the wet viable tissues.

However, there have been problems with bruising of the skin and particles bouncing off the skin surface.

A similar device is the Intraject, which is a development of the vaccine gun designed to deliver liquids through the skin without the use of needles.

TRANSDERMAL THERAPEUTIC SYSTEMS

Probably the most innovative practical step in the science of transdermal delivery in recent years has been the introduction into medicine of skin patches.

The original Transdermal Therapeutic System (TTS) or Transdermal (Drug) Delivery System (T(D)DS) was introduced as a device that would release drug to the skin at a controlled rate, well below the maximum the tissue can accept. Thus, the device, not the stratum corneum, would control the rate at which a drug diffuses through the skin, as the intended flux would be much lower than the maximum skin flux.

A TTS tries to provide systemic therapy in a more convenient and effective way than parenteral or oral therapy. The claimed advantages for the percutaneous over the oral route include:

1. Drug administration through the skin eliminates variables that influence gut absorption, such as changes in pH along the gastrointestinal tract, food and fluid intake, stomach emptying time and intestinal motility and transit time, and the presence of human and bacterial enzymes.
2. Drug enters the systemic circulation directly, eliminating the 'first-pass' effect of enzymes in the gut and the liver, the body's main metabolizing organ (but note that skin has its own enzymes and may thus metabolize some 5% of drug types).
3. Transdermal input may provide controlled, constant drug administration, displaying a single pharmacological effect. The continuity of input may permit the use of drugs with short half-lives and improve patient compliance.
4. Percutaneous administration could eliminate pulse entry into the circulation. Peaks in plasma concentrations often produce undesirable effects and troughs may be subtherapeutic.
5. The transdermal route can use drugs with a low therapeutic index.
6. Patches may be readily removed, although there is a reservoir effect and blood levels do not fall immediately to zero after TTS removal.

Transdermal systems usually contain potent drugs that should not irritate or sensitize the skin; they must be stable and have the correct physicochemical properties to partition into the stratum corneum and permeate to the vasculature.

Device design

Manufacturers design patches in a variety of ways, but for simplicity they may be categorized into one of two main types, the monolith (or matrix) or the rate-limiting membrane configurations. In considering these two designs, it is convenient initially to accept the original assumption that the skin under the patch operates as a perfect sink, even though no TTS produced to date works perfectly on this basis.

Monolith or matrix system

In these patches, the Higuchi square root of time law is usually obeyed. Equations 33.9–33.14 illustrate the relationships when the drug is dissolved in the matrix or exists as a suspension; Figure 33.13 illustrates release profiles, plotted both linearly and as square route functions of time. Figure 33.14 illustrates the fundamental construction for a suspension-type TTS. An occlusive backing layer protects the drug matrix, which comprises a suspension of drug in equilibrium with its saturated solution (maximum thermodynamic activity). An adhesive layer contains dissolved drug in equilibrium with that in the matrix, and attaches the patch to the skin. A protective strippable film is removed prior to application.

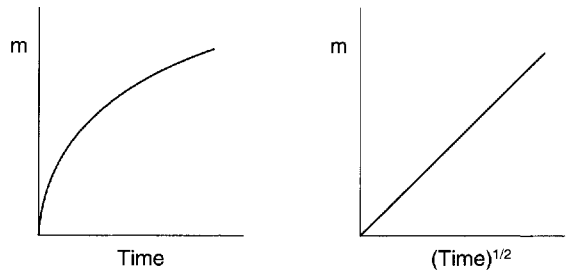


Fig. 33.13 Release profiles, plotted both linearly and as square route function of time, for matrix or monolith patches operating under Higuchi conditions (see Eqn 33.13).

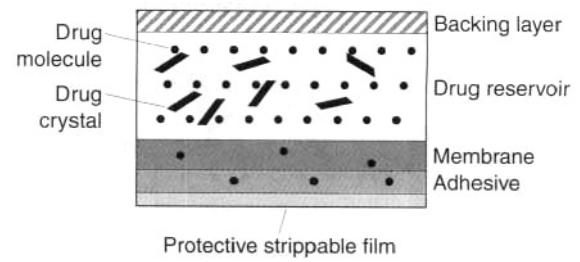


Fig. 33.14 Fundamental construction for a suspension type Transdermal Therapeutic System based on a matrix or monolith design (not to scale).

Rate-limiting membrane system

As these patches include a membrane, we might expect that the release profile would follow the simple Fickian conditions considered at the beginning of this chapter. Thus, at steady state we would expect the amount of drug released to the skin to be directly related to time (see Eqn 33.3 and Fig. 33.3). However, such a profile only follows when the membrane is initially free of drug. The lag time then represents the period during which the membrane equilibrates with drug ***after application to the sink*** (in this case the skin). In practice this does not happen, because the drug equilibrates in all patch components on storage, before the patient receives the patch.

Figure 33.15 illustrates the situation; a typical patch in this category consists of a backing layer, a reservoir containing the drug, the membrane, a skin adhesive and the protective film. On storage, the drug equilibrates into the membrane and adhesive. This portion of the drug more readily releases into the skin, as it does not have to permeate through all of the membrane. The result is to produce a so-called burst effect that leads to the type of plot illustrated in Figure 33.16. Such profiles may be confused with Higuchi plots, i.e. with matrix release plots, as illustrated in the first plot of Figure 33.13. An advantage is that the burst component can provide a quick-acting, priming dose of drug.

Future trends

It gradually became apparent to pharmaceutical scientists how difficult it was to formulate a TTS in

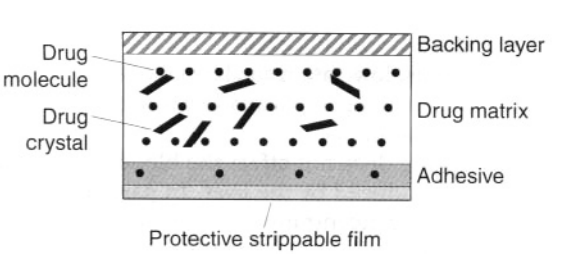


Fig. 33.15 Transdermal Therapeutic System based on a rate-limiting membrane design (not to scale).

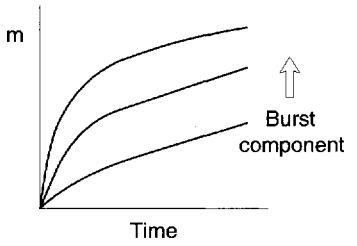


Fig. 33.16 Release profile from rate-limiting membrane patches; effect of increasing amount of drug partitioned into the membrane and adhesive on storage (which provides the 'burst' component).

such a way that the control of delivery remained within the patch and not in the very impervious skin of the patient. It is also expensive to manufacture complex, multilayered patches, particularly those containing membranes. There is thus a trend to concentrate on simple designs that are also thinner and therefore less obtrusive (cosmetically more acceptable). The result is a move towards the simple adhesive matrix patch illustrated in Figure 33.17.

Clinical patches

The following section discusses some examples of TTSs used in therapy. Besides their inherent clinical interest and value, each one exemplifies one or more biopharmaceutical features of importance in this mode of disease management.

Transdermal scopolamine (hyoscine)

The first patch brought to market delivered hyoscine through the thin, relatively permeable postauricular skin, via a rate-controlling membrane. The contact adhesive liberates priming drug to saturate horny layer binding sites and thus reduces the time before the clinical effect arises.

The TTS treats motion sickness and, apart from its value to cruise passengers, it was used by American astronauts. It can control the emetic side-effects of anticancer drugs, inhibit radiation sickness and

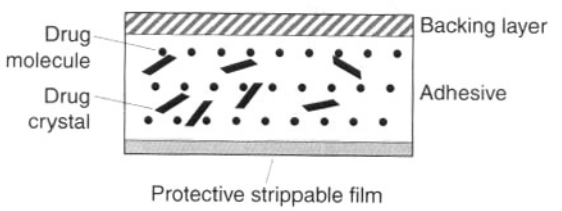


Fig. 33.17 Simple drug-in-adhesive patch (not to scale).

control vertigo. Two-thirds of patients still experience a dry mouth, but woodwind instrumentalists in orchestras have turned even this to advantage: they have found that application of the patch suppresses salivation.

This patch provides an excellent example of the original basic philosophy of TDDSs. In the usual dosage form of injections or tablets, hyoscine can produce side-effects such as excitement, confusion and hallucinations, which the controlled input from the TDDS eliminates. About 70% of the control resides in the patch; the variable nature of the patient's skin affects the remainder.

Transdermal nitroglycerin (glyceryl trinitrate, GTN)

A major use of TDDS technology is to deliver nitroglycerin, isosorbide dinitrate and isosorbide mononitrate for the treatment of angina, congestive heart failure and acute myocardial infarction. The nitroglycerin patch can also be valuable for the prophylactic treatment of phlebitis and extravasation secondary to venous cannulation.

The introduction of the first nitroglycerin device led to a vigorous debate as to the relative merits of patches, ointments and tablets. The main arguments revolved around whether or not the patch system delivered suboptimal drug levels to the blood; whether tolerance developed to the nitroglycerin; and if the TDDS or the patient's skin controlled the rate of GTN input. With regard to the last point, it was concluded that only about 50% or less of the control was inherent within the device.

Transdermal oestradiol

Several oestradiol systems treat women for menopausal symptoms or other results of oestrogen deficiency (hormone replacement therapy, HRT). The claimed benefits of transdermal delivery over oral therapy for such patches include:

- First-pass metabolism is avoided; therefore a lower dose is administered.
- Continuous, low-rate transdermal dosing maintains oestrogen plasma levels within physiological limits.
- Such dosing does not affect the blood levels of proteins produced by the liver. In particular, the angiotensin concentration remains the same and therefore the risk of hypertension is minimized. (Oral oestrogen therapy has been likened to hitting the liver with a sledgehammer!) However,

as liver enzymes are not stimulated, the beneficial effects on serum lipids are absent.
- Large increases in oestrone (the major metabolite of oestradiol) are avoided and the oestrone:oestradiol ratio restores to the premenopausal value.

Originally, the oestrogen patch was marketed only for women who had undergone a hysterectomy, because unopposed oestrogen (that which is given alone) can cause endometrial hyperplasia and increase the risk of endometrial cancer. Now women with an intact uterus can use combination patches that deliver oestradiol for 2 weeks, followed by a further 2 weeks of a progestin such as norethysterone acetate or levonorgestrel.

Patches are applied once or twice weekly and are effective in treating menopausal and post-menopausal symptoms such as flushing and vaginal atrophy. They also appear to be valuable in treating osteoporosis. Their contribution to reduction in the risk of cardiovascular disease remains to be clarified.

Transdermal clonidine

This formulation was originally introduced to treat hypertension. It was then reported to ameliorate some of the short-term symptoms associated with stopping smoking (craving, irritability, anxiety, restlessness and hunger).

After its introduction, clinical trials revealed a high incidence of topical irritation to clonidine (up to 50% of patients were affected). This finding emphasizes the importance of screening drug candidates for such side-effects before committing significant resources to a transdermal delivery programme.

Transdermal fentanyl

This opioid analgesic patch treats chronic intractable pain, such as that produced by cancer, over a 72-hour period. A TTS providing 25 μg of fentanyl per hour is approximately equivalent to the oral administration of 90 mg of morphine sulphate daily.

Transdermal nicotine

There are several such patches available and commercial competition is intense because of the huge worldwide market for smoking cessation therapy. Within the first year of introduction, total sales exceeded that for any other drug in history. However, as is common to most forms of antismoking treatment, subsequent yearly sales fell dramatically.

The TTSs provide an alternative to treatment via nicotine chewing gum, lozenges, sublingual tablets, nasal sprays and inhalers. The patches, designed to be worn for 16 or 24 hours, are available in different strengths so as to provide a weaning process. The concept is to try to match the plasma trough levels of nicotine produced by cigarette smoking, but not to mimic the 'hit' arising from inhalation. (A bolus of nicotine reaches the brain within seconds after inhaling from a cigarette – this is the main ***physiological*** component of addiction).

The nicotine patch can also maintain labour when there is danger of a premature delivery, and a patient with Parkinson's disease has claimed that the device had 'restored him to his pre-Parkinson's self'. Tourette's syndrome is a bizarre disorder characterized by jerky motions, rage and a propensity to utter obscenities: nicotine patches improved the effectiveness of drugs used in the treatment of afflicted children.

Nicotine is an ideal skin penetrant as it has a low molar mass, is liquid (low melting point), has a balanced partition coefficient and is miscible with oil and water (see Ideal properties for drug penetration, earlier in this chapter).

Transdermal testosterone

Hypogonadism (arising from pituitary or testicular disorder) afflicts 1 in 200 men, and testosterone deficiency may also arise from accidents and orchidectomy. The first patch developed was applied to the shaved scrotum, as scrotal skin is the most permeable skin site in males. With developments in formulation, more convenient areas, such as the back, abdomen, thigh or upper arm, may be used.

General conclusions on the usage of transdermal patches

- In recent years the formulation of transdermal patches has been a vibrant developmental area within the pharmaceutical industry.
- However, the original concept that the patch, not the skin, should control drug input to the patient has not been realized; no marketed patch fully controls the drug flux.
- There has been a move from a complex patch structure towards a simpler matrix formulation.
- Future progress will depend on:
  - correct choice of drug
  - synthesis of prodrugs
  - development of suitable penetration enhancers.

- There is a need to solve problems relating to, e.g.:
  - irritancy,
  - sensitization,
  - cutaneous metabolism,
  - localized body load of drug,
  - wearability of the patch for up to 7 days; the device must:
    - be thin, flexible and unobtrusive
    - adhere well when the patient sweats, exercises and bathes
    - not encourage microbial growth
    - not collect dirt around the periphery (where the adhesive contacts the environment).

FORMULATION OF DERMATOLOGICAL VEHICLES

People apply many skin preparations, ranging from powders, through semisolids to liquids. Formulators have often in the past developed such preparations for stability, compatibility and patient acceptability rather than considering the influences that the components may have on drug bioavailability. Modern formulation methods nowadays, however, have to concentrate on biopharmaceutical principles.

A valuable approach in designing a vehicle to produce optimum bioavailability is to use fundamental permeation theories, while remembering that the treatment regimen and the diseased skin usually violate the constraints of simple diffusion theory. However, with our present knowledge, such a formal approach may often limit the investigator to a single-phase system such as a polar gel; multiphase systems usually provide intractable theoretical problems. But physicians usually want a topical application to provide several therapeutic effects, as well as good absorption. These aims include anti-inflammatory efficacy in acute inflammation, symptomatic relief of pain and itch, protection from irritation, cleansing, and lubricant and emollient actions. A single-phase vehicle cannot readily achieve so much: complex multicomponent bases are necessary. Patients also tend to favour creams rather than gels or ointments.

This section considers the formulation of vehicles mainly in terms of ***unmedicated*** preparations. A good base must foster the remarkable recuperative capability of skin. For minor conditions it is often as important to select a vehicle that promotes healing and does no further damage, as it is to apply a therapeutic agent. A general rule is that for wet lesions the patient should use an aqueous dressing, and for dry skin a lipophilic base is best. Most vehicles are blended from one or more of three main components – aqueous solvents, powder and oil – together with thickening and emulsifying agents, buffers, antioxidants, preservatives, colours, propellants etc. We shall consider typical examples of the main types of skin preparation.

Dermatological formulations

Liquid preparations

Liquid preparations for external application include simple soaks or baths, applications, liniments, lotions, paints, varnishes, tinctures, and ear drops. A simple soak provides an active ingredient in aqueous solution or suspension, sometimes with water-miscible solvents. Gums and gelling agents may vary the consistency, from mobile liquids to stiff ringing gels. Bath additives such as Oilatum Emollient deposit a layer of liquid paraffin on the stratum corneum in an attempt to maintain its moisture content by occlusion. Applications may be liquid or viscous and often incorporate parasiticides, e.g. dicophane, benzyl benzoate, gamma benzene hexachloride and malathion. Liniments may be alcoholic or oily solutions or emulsions, which should not be applied to broken skin. Lotions are aqueous solutions or suspensions from which water evaporates to leave a thin uniform coating of powder. Evaporation cools and soothes the skin, making lotions valuable for treating acutely inflamed areas. Alcohol enhances the cooling effect and glycerol sticks the powder to the skin. Lotions may also be dilute emulsions, usually of the oil-in-water type. Paints, varnishes and tinctures present solutions of active ingredients in volatile solvents such as water, industrial methylated spirits, acetone or ether. Ear drops are often aqueous solutions, although glycerol and alcohol may also be used.

Gels (jellies)

Gels are two-component semisolid systems rich in liquid. Their one characteristic feature is the presence of a continuous structure providing solid-like properties. In a typical polar gel, a natural or synthetic polymer builds a three-dimensional matrix throughout a hydrophilic liquid. Typical polymers used include the natural gums tragacanth, carrageenan, pectin, agar and alginic acid; semisynthetic materials such as methylcellulose, hydroxyethylcellulose, hydroxypropylmethylcellulose and carboxymethylcellulose; and the synthetic polymer Carbopol

(carbomer). Certain clays such as bentonite, Veegum and Laponite may be also used. Provided that the drug does not bind to the polymer or clay, such gels release medicaments well; the pores allow relatively free diffusion of molecules that are not too large.

Powders

Dusting powders for application to skin folds are finely divided (impalpable) insoluble powders which dry, protect and lubricate, e.g. talc, zinc oxide, starch and kaolin. Dusting powders should not contain boric acid, as abraded skin may absorb it in toxic amounts.

Ointments

Ointments are greasy, semisolid preparations, often anhydrous and containing dissolved or dispersed medicaments.

Hydrocarbon bases These usually consist of soft paraffin or mixtures with hard paraffin. Paraffins form a greasy film on the skin, inhibiting moisture loss and improving hydration of the horny layer in dry scaly conditions. This hydration is also a main reason why ointments are so effective in encouraging percutaneous absorption of a drug.

The Plastibases are a series of hydrocarbons containing polyethylene, which forms a structural matrix in systems which are fluid at the molecular scale but are typical dermatological semisolids. They are soft, smooth, homogenous, neutral, colourless, odourless, non-irritating, non-sensitizing, extremely stable vehicles. Plastibases are compatible with most medicaments and they maintain their consistency even at high concentrations of solids and under extremes of temperature. The bases apply easily and spread readily, adhere to the skin, imparting a velvety, non-greasy feel, and can readily be removed.

Soap-based greases may be produced by, for example, incorporating aluminium stearate in a heavy mineral oil. A random arrangement of metallic soap fibres weaves throughout the oil, producing a product which changes its consistency only slightly when heated; the base readily incorporates drugs and the addition of lanolin permits the absorption of a little water.

Fats and fixed-oil bases Dermatological vehicles have frequently contained fixed oils of vegetable origin, consisting essentially of the mono-, di- and triglycerides of mixtures of saturated and unsaturated fatty acids. The most common oils include peanut, sesame, olive, cottonseed, almond, arachis, maize and persic. Such oils can decompose on exposure to air, light and high temperatures, and may turn rancid. Trace metal contaminants catalyse oxidative reactions which the formulator combats with antioxidants, such as butylated hydroxytoluene, butylated hydroxyanisole or propyl gallate, or with chelating agents such as the salts of ethylenediaminetetraacetic acid (EDTA). However, antioxidants may be incompatible with the drug or they may sensitize some patients.

Silicones Dimethicones, or dimethyl polysiloxanes, have properties similar to hydrocarbon bases. They are water repellent with a low surface tension and are incorporated into barrier creams to protect the skin against water-soluble irritants.

Absorption bases Absorption bases soak up water to form water-in-oil emulsions while retaining their semisolid consistencies. Generally, they are anhydrous vehicles composed of a hydrocarbon base and a miscible substance with polar groups that functions as a water-in-oil emulsifier, e.g. lanolin, lanolin isolates, cholesterol, lanosterol and other sterols, acetylated sterols, or the partial esters of polyhydric alcohols such as sorbitan monostearate or mono-oleate. Bases such as Wool Alcohols Ointment BP and Simple Ointment BP deposit a greasy film on the skin, similar to a hydrocarbon base, but they suppress less the transepidermal water loss. However, they may hydrate the stratum corneum by applying a water-in-oil emulsion, thereby prolonging the time during which the horny layer can absorb moisture. Some individuals are sensitive to lanolin. This may be important because the sensitization occurs unexpectedly and physicians frequently overlook it, especially in atopic patients, who may apply large quantities of lanolin-containing emollients for protracted periods. Modern purified lanolin preparations are less sensitizing.

Emulsifying bases These essentially anhydrous bases contain oil-in-water emulsifying agents which make them miscible with water and so washable or 'self-emulsifying'. There are three types, depending on the ionic nature of the water-soluble emulsifying agents: anionic (e.g. Emulsifying Ointment), cationic (e.g. Cetrimide Emulsifying Ointment) and non-ionic (e.g. Cetomacrogol Emulsifying Ointment). Because they contain surfactants, emulsifying bases may help to bring the medicament into more intimate contact with the skin. The bases mix with aqueous secretions and readily wash off the skin; thus they are useful for scalp treatments.

Water-soluble bases Formulators prepare water-soluble bases from mixtures of high and low molecular weight polyethylene glycols (macrogols, Carbowaxes). Suitable combinations provide products with an ointment-like consistency, which

soften or melt on skin application. They are non-occlusive, mix readily with skin exudates and do not stain sheets or clothing; washing quickly removes any residue. The macrogols do not hydrolyse, deteriorate, support mould growth or irritate the skin. Examples include Macrogol Ointment and Polyethylene Glycol Ointment; because they are water-soluble, they will not take up more than 8% of an aqueous solution before losing their desirable physicochemical characteristics. To enable a base to incorporate more water stearyl alcohol can be substituted for some of the macrogol component. Macrogol bases are used with local anaesthetics such as lignocaine, but they are incompatible with many chemicals, including phenols, iodine, potassium iodide, sorbitol, tannic acid, and the salts of silver, mercury and bismuth. The bases diminish the antimicrobial activity of quaternary ammonium compounds and methyl and propyl parahydroxybenzoates, and rapidly inactivate bacitracin and penicillin.

Creams

Creams are semisolid emulsions for external application. Oil-in-water emulsions are most useful as water-washable bases, whereas water-in-oil emulsions are emollient and cleansing. Patients often prefer a w/o cream to an ointment because the cream spreads more readily, is less greasy, and the evaporating water soothes the inflamed tissue. O/w creams ('vanishing' creams) rub into the skin; the continuous phase evaporates and increases the concentration of a water-soluble drug in the adhering film. The concentration gradient for drug across the stratum corneum therefore increases, promoting percutaneous absorption. To minimize drug precipitation, a formulator may include a less volatile, water-miscible cosolvent. An o/w cream is non-occlusive because it does not deposit a continuous film of water-impervious liquid. However, such a cream can deposit lipids and other moisturizers on and into the stratum corneum and so restore the tissue's hydration ability, i.e. the preparation has emollient properties.

It is difficult to predict the role that an emulsion plays in percutaneous absorption. This is because, added to all the physiological and physicochemical considerations already discussed, the following must also be considered:

- Partitioning of the medicament between the emulsion phases;
- The addition of preservatives;
- Determination of a true viscosity for the diffusing molecules in the vehicle;
- The possibility of phase inversion or cracking of the emulsion when applied to the skin.

Drug may also be trapped in the micelles and the gel and liquid crystalline phases present in the continuous phase. Emulsions are complex systems and so all medicaments must be considered individually with respect to emulsion design. There are few worthwhile formulation guidelines additional to the principles already discussed.

Pastes

Pastes are ointments containing as much as 50% powder dispersed in a fatty base. They may be useful for absorbing noxious chemicals in babies, such as the ammonia that bacteria release from urine. Because of their consistency, pastes localize the action of an irritant or staining material, such as dithranol or coal tar. They are less greasy than ointments because the powder absorbs some of the fluid hydrocarbons. Pastes lay down a thick, unbroken, relatively impermeable film that can be opaque and act as an efficient sun filter. Skiers use such formulations on the face to minimize windburn (excessive dehydration) and to block out the sun's rays.

Aerosols

Aerosols may function as drug delivery systems for solutions, suspensions, powders, semisolids and emulsions.

Solution aerosols are simple products with the drug dissolved in a propellant or a propellant/solvent mixture. Typical agents incorporated are steroids, antibiotics and astringents. The powder aerosol methodology is useful for difficult soluble compounds such as steroids and antibiotics. Semisolid preparations, such as ointments and creams, may be prepared in a flexible bag with compressed nitrogen used for expulsion instead of a volatile propellant. Emulsion systems produce foams that may be aqueous or non-aqueous and stable or quick-breaking. Medicinal stable foams are aqueous formulations used, for example, for preoperative shaving and for contraception. The stable foam, which is similar to a medicated shaving cream, varies in stability depending on the surfactant, solvent and propellant used.

Cosmetic or aesthetic criteria for dermatological formulations

However well designed a topical vehicle for maximum drug bioavailability is, the preparation

must still be acceptable to the patient. A product that is poor in appearance may lead to non-compliance. Patients generally prefer a formulation which is easy to transfer from the container, spreads readily and smoothly, leaves no detectable residue, and adheres to the treated area without being tacky or difficult to remove. The dosage of such preparations is typically in the range of 1–5 mg/cm^2 of skin.

Stiff pastes may be hard to rub into the skin or to apply evenly; application to damaged areas may therefore be painful. However, a thick layer of material can occlude the tissue or protect it from mechanical, chemical or light damage. Ointments and pastes do this, and the viscous drag imposed on application may dislodge scales, dead tissue and the remnants of previous doses. The medicament then makes intimate contact with the diseased site. A stiff preparation also helps to delineate the area of treatment.

The sensations of greasiness and tackiness arise from those constituents that form the skin film. For creams, stearic acid and cetyl alcohol produce non-tacky films. Formulations that include synthetic or natural gums should use the minimum amount, as the polymers tend to leave a tacky coating on the skin.

Insoluble solids leave an opaque layer that often appears powdery or crusty. However, as therapy requires such solids in lotions and pastes, the formulator can do little to vary the film's nature and patients accept the residue as part of the treatment.

Physicochemical criteria for dermatological formulations

The developer of dosage forms must note the physical and chemical behaviour of the drug and the dosage form during preformulation studies, throughout bench-scale work, pilot studies and batch processing, at the manufacturing level, and during storage and use of a product. Some general factors that a pharmaceutical scientist evaluates for a new semisolid during developmental studies and storage include:

- Stability of the active ingredients
- Stability of the adjuvants
- Rheological properties – consistency, viscoelasticity, extrudability
- Loss of volatiles, including water
- Phase changes – inhomogeneity, bleeding, cracking
- Particle size distribution of dispersed phase
- Apparent pH
- Particulate contamination.

The first difficulty arises in assessing the chemical stabilities of the drug (in its complex vehicle) and the adjuvants. A general method establishes a shelf-life by using an accelerated stability test at elevated temperatures and the Arrhenius relationship. However, for a multiphase system such as a cream, heat may change the phase distribution and may even crack the emulsion. Thus, the investigator may have to assess the preparation for a long time at the storage temperature. Because of vehicle complexity, it may be difficult to separate the labile components for analysis.

Many dermatologicals contain volatile solvents, and batches may lose some solvent through the walls of plastic containers or through faulty seams or ill-fitting caps.

Heterogeneous systems may suffer phase changes when stored incorrectly. Emulsions may crack and cream, suspensions can agglomerate and cake, and ointments and gels may 'bleed' as their matrices contract and squeeze out constituents. High temperatures can produce or accelerate such adjustments. Multipoint rheological assessments can readily quantify structural changes in colloidal systems; viscoelastic determinations, such as creep and oscillation, are also valuable.

For suspensions and emulsions, a particle size analysis may often detect a potentially unstable formulation long before any other parameter changes markedly. Emulsion globules may grow through coalescence as gel networks break down on storage; crystals may enlarge or change their habit, or revert to a more stable, less active polymorphic form. Such alterations in crystal form may affect the therapeutic activity of the formulation.

The apparent pH of a topical product may change on storage. Although pH measurements of complex vehicles have no fundamental meaning, investigators sometimes use a pH electrode to monitor formulations as they age.

It can be difficult to manufacture creams and ointments completely free from foreign particles. Aluminium and tin tubes may contaminate a topical with 'flashings' – metal slivers and shavings formed during container fabrication. Their presence is particularly undesirable in ophthalmic ointments, and various pharmacopoeial tests limit the extent of such contamination. Plastic tubes are now generally used.

In addition to instrumental tests, the pharmaceutical scientist should note any qualitative changes during product storage. The colour may vary, e.g. natural fats, oils and lanolin brown as they oxidize, becoming rancid with a disagreeable odour. The texture may alter as phase relationships vary.

Microbial contamination and preservation: rancidity and antioxidants

Topical bases often contain aqueous and oily phases, together with carbohydrates and even proteins, and so bacteria and fungi readily attack them. Microbial growth spoils the formulation and is a potential toxic hazard and a source of infection. Conditions that lower immunity, such as injury, debilitating diseases or drug therapy, may encourage organisms that are usually not highly infectious to infect the host, i.e. to become opportunistic pathogens. In 1969, 33 samples of 169 cosmetics and topical drugs surveyed were microbially contaminated, half with Gram-negative organisms which were a health hazard. In the mid-1960s, an outbreak of serious eye infection was traced to an antibiotic ophthalmic ointment contaminated with *Pseudomonas*.

There are many potential sources of microbial contamination. It can occur in raw material and in the manufacturing water, in processing and filling equipment, in packing material, if there is poor plant hygiene or an unclean environment, and if plant operatives fail to comply with good manufacturing procedures.

Because of the complexity of dermatological vehicles and their manufacturing processes there exists no universal preservative, although we can summarize the essential requirements for selecting a material to preserve a specific formulation. The additive must be compatible with all ingredients; it should be stable to heat, prolonged storage and product use conditions; and it must be non-irritant, non-toxic and non-sensitizing to human tissue.

Many prototype pharmaceutical preparations could deteriorate on storage because some components oxidize when oxygen is present. This decomposition can be particularly troublesome in emulsions, because emulsification may introduce air into the product and because of the high interfacial contact area between the phases.

The ideal antioxidant would possess the following properties:

- Effective at low concentrations;
- It and its decomposition products should be non-toxic, non-irritant, non-sensitizing, odourless and colourless;
- Stable and effective over a wide pH range;
- Neutral – should not react chemically with other ingredients;
- Non-volatile.

PROTOCOL FOR DESIGNING, DEVELOPING AND TESTING A DERMATOLOGICAL FORMULATION

Below are listed steps that may help a pharmaceutical scientist to design a satisfactory formulation. The treatment, although condensed, is useful for providing a checklist to control the development programme.

1. Identify the disease or condition to be treated.
2. Determine the site for drug action – skin surface, stratum corneum, viable epidermis, dermis, appendages or systemic circulation. Consider the body region, e.g. scalp, trunk, feet, nails etc.
3. Note the receptor site within the target area (this may be unknown).
4. Estimate the condition of the average patient's skin – thickened (e.g. ichthyosis), broken and inflamed (e.g. acute eczema), pilosebaceous unit blocked (acne), etc. Remember that successful treatment may rapidly change the condition of the skin. For example, a weeping, wet skin without an intact horny layer may heal quickly to produce a few cell layers with a dry surface.
5. Choose the best drug or prodrug for the disorder; consider its pharmacological and pharmacokinetic profiles, toxicity, sensitizing potential, stability, susceptibility to skin enzyme metabolism and physicochemical properties (particularly the diffusion coefficient and partition coefficient relevant to the horny layer).
6. Evaluate the optimum kinetics for drug delivery to the target site. Consider pulsed or steady-state treatment, amount and strength of dosage form and frequency of application.
7. In the light of points 1–6 above, select the type of formulation needed, e.g. cream, ointment, aerosol, delivery device.
8. Decide whether and where there is a rate-limiting step in the treatment, e.g. solely within the vehicle, permeation across the stratum corneum, or clearance from the viable tissues. Concentrate on maximizing this rate.
9. Choose vehicle ingredients that are stable, compatible, and cosmetically and therapeutically acceptable. Be aware that these adjuvants may have their own therapeutic effects, e.g. occlusive vehicles moisturize the skin.
10. If the intention is to promote drug penetration, optimize the formulation to the maximum chemical potential of the drug. Remember that

vehicles often change after application as components evaporate or penetrate the skin and secretions mix with the formulation.

11. If the drug is a poor penetrant consider using a penetration enhancer, but remember that a new enhancer will need a full toxicological screen. Regulatory authorities are particularly cautious about the use of such promoters.
12. Perform in vitro tests with trial formulations using a simple synthetic membrane (or no membrane) and a suitable sink; ensure that the drug releases readily from the vehicle.
13. Repeat 12, preferably with human skin, to monitor permeation. Such experiments may include a steady-state design and a scheme that mimics clinical use (the so-called finite dose design).
14. Conduct in vivo studies in animals and human volunteers to check for efficacy, safety and acceptability; determine the pharmacokinetic profile and the topical bioavailability.
15. Do clinical trials.
16. Throughout the programme, review the physicochemical behaviour and stability of the dosage form and package during preformulation studies, scale-up procedures, manufacture, storage and use.

BIBLIOGRAPHY

Barry, B.W. (1983) *Dermatological Formulations: Percutaneous Absorption*. Marcel Dekker, New York. (Provides additional fundamental information and general background.)

Bronaugh, R.L. and Maibach, H.I. (eds) (1999) *Percutaneous Absorption*, 3rd edn. Marcel Dekker, New York. (A modern multiauthored text with many relevant chapters.)

Chien, Y.W. (1992) *Novel Drug Delivery Systems*. 2nd edn, Chapter 7. Marcel Dekker, New York. (First two chapters provide background on rate-controlled drug delivery.)

Hsieh, D.S. (ed) (1994) *Drug Permeation Enhancement*. Marcel Dekker, New York. (Parts I, II and III deal with drug enhancement through skin.)

Potts, R.O. and Guy, R.H. (eds) (1997) *Mechanisms of Transdermal Drug Delivery*. Marcel Dekker, New York. (Provides information on biophysical methods for determining molecular transport across skin.)

Roberts, M.S. and Walters, K.A. (eds) (1998) *Dermal Absorption and Toxicity Assessment*. Marcel Dekker, New York. (A substantial text with much relevant information.)

Smith, E.W. and Mailbach, H.I. (eds) (1995) *Percutaneous Penetration Enhancers*. CRC Press, Boca Raton. (Chapters deal with chemical and physical enhancement methods and theories, together with analytical techniques.)

Williams, A.C. and Barry, B.W. (1999) Raman spectroscopy. In: *Percutaneous Absorption* 3rd edn. (Bronaugh, R.L. and Maibach, H.I., eds), Marcel Dekker, New York, pp 499–514.

34
Rectal and vaginal drug delivery

Josef Tukker

CHAPTER CONTENTS

RECTAL DRUG DELIVERY

Introduction

The administration of drugs by routes other than the oral one has to be considered in several circumstances and for a great many varying reasons. Arguments for choosing the rectal route for drug administration include:

1. The patient is not able to make use of the oral route. This may be the case when the patient have a problem with their gastrointestinal tract, is nauseous or is postoperative (when they may be unconscious or not able to ingest a drug orally). Furthermore, several categories of patients, i.e. the very young, the very old or the mentally disturbed, may more easily use the rectal than the oral route.
2. The drug under consideration is less suited for oral administration. This may be so in cases where oral intake results in gastrointestinal side-effects; also, the drug may be insufficiently stable at the pH of the GI tract, or susceptible to enzymatic attack in the GI tract or during the first passage of the liver after absorption. Also, drugs with an unacceptable taste can be administered rectally without this inconvenience to the patient. The formulation into suppositories of certain drugs that are candidates for abuse, as in suicide, has also been considered.

Besides these apparent advantages, the rectal route also has several drawbacks. Depending on tradition, there are strong feelings of aversion in certain countries, such as the UK and the USA, to rectal administration of drugs, whereas there is complete acceptance on the continent and in Eastern Europe. More rational points in this respect are the slow and sometimes incomplete absorption that has been

reported, and the considerable inter- and intrasubject variation. The development of proctitis has also been reported. There are also problems with the large-scale production of suppositories and the achievement of a suitable shelf-life (the latter demanding stringent storage conditions).

It can thus be concluded that rectal administration should certainly not be the route of first choice, but can in certain circumstances be of great advantage to the patient. The rectal route is used in many different therapies, intended either for local or for systemic effect. ***Local effect*** is desired in the case of pain and itching, mostly due to the occurrence of haemorrhoids. Locally active drugs which are used include astringents, antiseptics, local anaesthetics, vasoconstrictors, anti-inflammatory compounds and soothing and protective agents. Also some laxatives fall into this category. For the attainment of a ***systemic effect*** all orally given drugs can be used and many are, bearing in mind the limitations discussed above. Antiasthmatic, antirheumatic and analgesic drugs are very much used for this purpose.

Anatomy and physiology of the rectum

Rectal dosage forms are introduced into the body through the anus and are thus brought into the most caudal part of the GI tract, i.e. the rectum. Anatomically the rectum is part of the colon, forming the last 150–200 mm of the GI tract.

The rectum can be subdivided into the anal canal and the ampulla, the latter forming approximately 80% of the organ. It is separated from the outside world by a circular muscle, the anus. The rectum can be considered as a hollow organ with a relatively flat wall surface, without villi and with only three major folds, the rectal valves. The rectal wall is formed by an epithelium which is one cell layer thick, and is composed of cylindrical cells and goblet cells which secrete mucus. A diagram of part of the rectal wall and the rectum's venous drainage is shown in Figure 34.1.

The total volume of mucus is estimated as approximately 3 mL, spread over a total surface area of approximately 300 cm^2. The pH of the mucous layer is reported as approximately 7.5. Furthermore, there seems to be little buffer capacity. This point will be discussed later in relation to absorption.

Under normal circumstances the rectum is empty; filling provokes a defecation reflex, which is under voluntary control. Data comparing drug absorption from freshly prepared and aged, more viscous suppositories suggest that there is enough motility to provoke the spreading even of rather viscous suppositories.

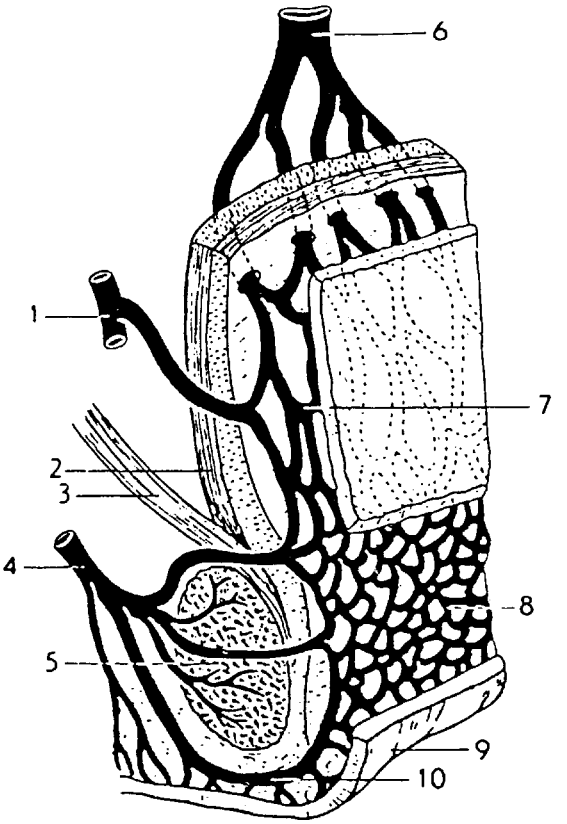


Fig. 34.1 Venous drainage of the human rectum (after Tondury, 1981): 1 middle rectal vein; 2 tunica muscularis: stratum longitudinale; 3 m. levator ani; 4 inferior rectal vein; 5 m. sphincter ani externus; 6 superior rectal vein; 7 and 8 plexus venosus rectalis (submucosus); 9 skin; 10 v. marginalis.

Absorption of drugs from the rectum

Blood supply, especially venous drainage, is important for the understanding of drug absorption. As can be seen from Figure 34.1, there are three separate veins. The lower and middle haemorrhoidal veins drain directly into the general circulation; the upper one drains into the portal vein, which flows to the liver. This means that drug molecules can enter the general circulation directly or by passing through the strongly metabolizing liver. In the latter case only a proportion of the drug molecules (if they are of the high clearance type) will enter the general circulation intact. Thus the bioavailability may be less than 100%. Compared to the small intestine this situation is still more favourable. Recent investigations have shown that avoiding the first passage through the liver is possible, but the extent of this effect cannot be generalized, as it will depend on the actual part of the rectum through which the drug is absorbed.

Thus keeping the drug in the lower part of the rectum would be advisable.

The insertion of a suppository into the rectum results in a chain of effects leading to the bioavailability of the drug. This is represented in a simplified scheme in Figure 34.2.

Depending on the character of its vehicle (see later) a suppository will either dissolve in the rectal fluid or melt on the mucous layer. Because the volume of rectal fluid is so small, dissolution of the complete vehicle will be difficult and require extra water. Owing to osmotic effects (of the dissolving vehicle) water is attracted, with the unpleasant consequence of a painful sensation for the patient. Independent of the vehicle type, drugs that are dissolved in the suppository will diffuse out towards the rectal membranes. Suspended drugs will first have to leave the vehicle (if it is water immiscible) under the influence of either gravity or motility movements, and can then start dissolving in the rectal fluid. The dissolved drug molecules will have to diffuse through the mucous layer and then into and through the epithelium forming the rectal wall.

The process of absorption will be a passive diffusion process, as it is throughout the whole GI tract for almost all drugs; active transport processes, as shown in the upper regions of the GI tract, have not been shown to be present in the rectal area.

For a generalized discussion on drug absorption the reader is referred to Part Three of this book. However, some specific points concerning rectal absorption will be discussed here. Table 34.1 gives a survey of the physiological factors in rectal absorption.

The quantity of fluid available for drug dissolution is very small (approximately 3 mL, spread in a layer approximately 100 μm thick over the organ). Only under non-physiological circumstances is this volume enlarged, e.g. by osmotic attraction by water-soluble vehicles or by diarrhoea. Thus the dissolution of slightly soluble substances, for example phenytoin, can easily be the slowest step in the absorptive process.

The properties of the rectal fluid, such as composition, viscosity and surface tension, are essentially unknown and have to be estimated from data available for other parts of the GI tract. The pH and the buffer capacity of the rectum were mentioned earlier in this chapter. The rectum is usually empty, except temporarily when faecal matter arrives from higher parts of the colon. This material is either expelled or

Table 34.1 Physiological factors affecting absorption from the rectum

Quantity of fluid available
Properties of rectal mucus
Contents of the rectum
Motility of the rectal wall

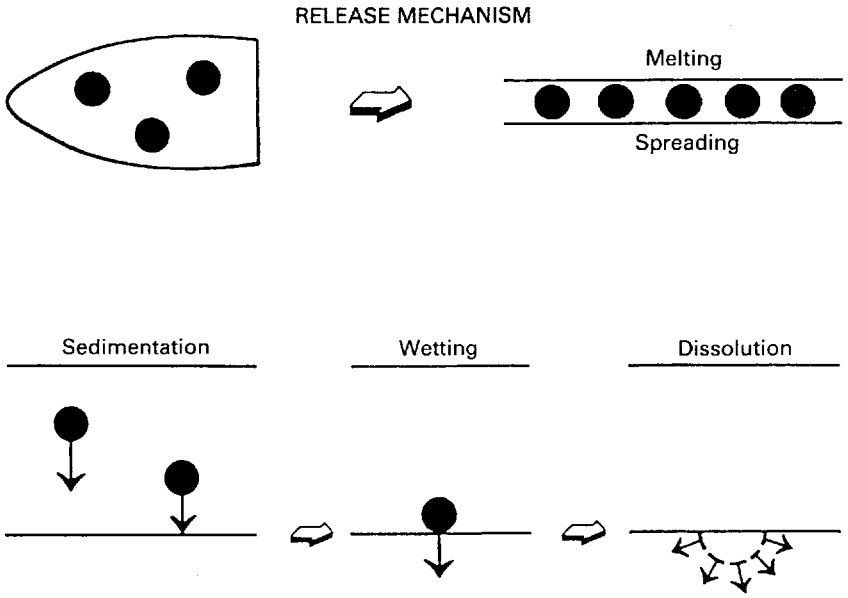


Fig. 34.2 Release process of a drug from a suspension suppository.

transported back into the colon, depending on the voluntary control exhibited on the anal sphincter. The rectal wall may exert a pressure on a suppository present in the lumen by two distinct mechanisms. First, the abdominal organs may simply press on to the rectum, especially when the body is upright. This may stimulate spreading and thus promote absorption. The second source of pressure is the motility of the muscles of the rectal wall, which may originate from the normally occurring colonic motor complexes. These are waves of contractions running over the wall of the colon in a caudal direction and are associated with the presence of food residues in the colon.

In contrast with the upper part of the GI tract, in the rectum no esterase or peptidase activity is present, resulting in a much greater stability of peptide-like drugs. Administration of these compounds using the rectal or vaginal route has been satisfactory, but only if absorption enhancers such as surfactants were used concomitantly. All kinds of surfactants seem to do the promoting work, of which polyoxyethylene lauryl alcohol ether appears to be the most powerful. One major drawback of these enhancers, however, is the irritation of the rectal mucosa in the long term; less irritating enhancers are needed to explore this interesting area in greater depth.

Formulation of suppositories

Suppositories are used mainly for the administration of drugs via the rectal route, but not exclusively. Application via other routes, such as the vagina, is less common but of distinct use in the treatment of locally occurring infections. Other suppository-type products (bougies) used through other body orifices, e.g. the ear, nose and urethra, are very uncommon and are not discussed here. Alternative dosage forms for the rectal and or vaginal route are tablets, capsules, ointments and enemas. These will be discussed later, and concentrate first on rectal suppositories. Suppositories are formulated in different shapes and sizes (usually 1–4 g). Their drug content varies widely, from less than 0.1% up to almost 40%. A more detailed description, together with the methods of preparation can be found in *Pharmaceutical Practice* (Winfield and Richards 1998). Generally the suppository consists of a vehicle in which the drug is incorporated, and in some cases additives are coformulated.

The vehicle (suppository base)

There are two main classes of vehicles in use, the glyceride-type fatty bases and the water-soluble ones. Although the ideal vehicle has not been found, the large variety of bases that are available enables a well considered choice for every drug that has to be formulated as a suppository.

Choosing the optimum base requires a lot of practical experience and at present this can only partly be guided by scientifically sound data. Much remains to be learned here. However, some general guidelines can be given.

Requirements of the vehicle There is no doubt that a suppository should either melt after insertion in the body or dissolve in (and mix with) the available volume of rectal fluid. For fatty bases this means a melting range lower than approximately 37°C (one must be aware of the fact that the body temperature might be as low as 36°C at night). The melting range should be small enough to give rapid solidification after preparation, thereby preventing the separation of suspended, especially high-density, drug particles and agglomeration. When the solidification rate is high this may result in fissures, especially when rapid cooling is applied. On the other hand, the melting range should be large enough to permit easy preparation, which on an industrial scale may take a considerable length of time.

During solidification a suppository should exhibit enough volume contraction to permit removal from the mould or plastic former. The viscosity of the molten base plays an important role both from a technological and from a biopharmaceutical point of view. During preparation the viscosity determines the flow into the moulds, but also the separation of drug particles. Clearly a compromise has to be found here. During and after melting in the rectal cavity the suppository mass is forced to spread over the absorbing surface, the rate of which may be determined partly by its viscosity. Drug particles that have to be transported through the base to the interface with the rectal fluid, and have to pass this interface to be released, will evidently also see viscosity as a determining factor in their journey.

A good suppository base should further be chemically and physically stable during storage as a bulk product, and after preparation into a suppository. It should have no incompatibility with drug molecules and should permit an optimal release of the drug it contains.

Clearly this list of requirements cannot always be completely fulfilled, and often an acceptable compromise is the best that one can expect.

Fatty vehicles The fatty vehicles in use nowadays are almost exclusively semi- or fully synthetic ones. Cocoa butter is no longer used because of its many disadvantages, such as its well known polymorphic

behaviour, its insufficient contraction at cooling, low softening point, chemical instability, poor water-absorptive power and its price.

The semisynthetic type of fatty vehicles (sometimes termed ***adeps solidus***) have few or none of the problems mentioned above. A comparison is made in Table 34.2.

The general composition of both types is mixed triglycerides with C_{12}–C_{18} acids. In the semisynthetic vehicles these acids are saturated, whereas cocoa butter contains a considerable amount of the unsaturated oleic acid (see iodine number in the table, which for reducing drugs should be <0.5). The presence of oleic acid is almost solely responsible for the special properties of this vehicle. The melting range of the (semi)synthetic bases is usually approximately 3°C and higher than that of cocoa butter; the acid content is lower (mostly <0.5), which is one of the reasons that the ageing of aminophylline suppositories is slower when (semi)synthetic vehicles are used. The hydroxyl number in the table refers directly to the amount of mono- and diglycerides present in the fatty base. A high number means that the power to absorb water is high. This may lead to an increased rate of decomposition for drugs that are easily hydrolysed, such as acetylsalicylic acid. It should be realized that this capacity could lead to the formation of a w/o emulsion in the rectum, which is generally to be avoided because of its very low drug release rate. An advantage of a high hydroxyl number is the larger melting and solidifying range, which permits easier manufacture.

Water-soluble vehicles Water-soluble (or miscible) vehicles are much less used, for reasons to be discussed below. They comprise the classic glycerol–gelatin or soap bases, which are used exclusively for laxative purposes or in vaginal therapy.

The macrogols are also used. They consist of mixtures of polyethylene glycols of different molecular weight. The melting point is well over body temperature, which means that they mix with the rectal fluid. For true dissolution the available volume of rectal fluid (1–3 mL) is too small. Because of their high melting point they are especially suited for application in tropical climates, but several disadvantages must be considered. They are hygroscopic and therefore attract water, resulting in a painful sensation for the patient. The incorporation of at least 20% water and moistening before insertion can help to reduce this problem. A considerable number of incompatibilities with various drugs (e.g. phenols, sulphonamides) has been reported. Owing to the solubilizing character of this base (low dielectric constant) drugs may tend to remain in the base and release may be slow.

Choice of vehicle A summary of the points that are important for the choice of a suppository base is given in Table 34.3. The parameters mentioned are evidently not independent of each other, and one interesting parameter can be added to this list, i.e. the volume of the suppository. Usually suppositories for adults are 2 mL and those for children 1 mL. It has been suggested that the larger volume may provoke a reaction in the rectal wall, thus helping to spread the melt over a larger area. Indeed the increase in volume of, for example, paracetamol suppositories, resulted in faster and more complete drug absorption.

The drug

Table 34.4 lists the factors related to the drug substance that are of possible consequence for the quality of suppositories.

Drug solubility in vehicle The drug solubility in the vehicle is of particular interest from the biopharmaceutical point of view, as it directly determines the

Table 34.2 Some properties of fatty suppository bases

	Melting range (°C)	Acid content	Hydroxyl number	Iodine number
Cocoa butter	31–34	<5	0	34–38
Adeps solidus	33–37.5	<2	<5–30	<3

Table 34.3 Formulation parameters of suppository bases

Composition
Melting behaviour
Rheological properties

Table 34.4 Drug-substance related factors

Solubility in water and vehicle
Surface properties
Particle size
Amount
pK_a

type of product, i.e. solution or suspension suppository. The drug solubility in the rectal fluid determines the maximum attainable concentration and thus the driving force for absorption. When a drug has a high vehicle to water partition coefficient it is likely to be in solution to an appreciable extent (or completely) in the vehicle. This generally means that the tendency to leave the vehicle will be small and so the release rate into the rectal fluid will be low. This is obviously unfavourable for rapid absorption. On the other hand, a certain lipid solubility is required for penetration through the rectal membranes (see above, under Absorption of drugs from the rectum). The optimal balance between these two requirements is usually found using the rules listed in Table 34.5. This table assumes that the release from the dosage form is considered as the rate-limiting step. Thus the tendency to remain in the base should be lowered as much as possible (rules 1 and 2). When the solubility in fat and water are both low no definite rule can be given. It may well be that the dissolution rate will become the controlling step, and thus it seems advisable to use micronized drug particles.

It should be stated as a general rule that emulsion-type suppositories (w/o) are strongly discouraged. The transfer of drug molecules present in dissolved state in the inner phase will be very slow, and so the absorption will be very much retarded.

It seems logical, therefore, that the first choice of a formulation would be a readily water-soluble form of the drug dispersed in a fatty base. This lays special emphasis on the water solubility of drugs and the methods to improve this. The role of pK_a in this respect should also be considered. For a detailed discussion of these points, see Chapters 2, 3, 8 and 17.

Surface properties The surface properties of drug particles are also important, as these particles will be transferred from one phase to another (see Fig. 34.2). This happens first when the drug is brought into contact with the vehicle and air has to be displaced from its surface. When this is not achieved particles may form agglomerates. This adversely affects final content uniformity by creating an increased tendency to separate. If wetting by the vehicle has taken place displacement by rectal fluid will be required to let the drug go into solution, which is the prerequisite for absorption. This is the underlying reason why people have tried the addition of surfactants to their formulation (see below).

Table 34.5 Drug solubility and suppository formulation

Solubility in		Choice of base
Fat	Water	
Low	High	Fatty base (rule 1)
High	Low	Aqueous base (rule 2)
Low	Low	Indeterminate

Particle size The particle size of the drug is an important parameter, both technologically and biopharmaceutically. To prevent undue sedimentation during or after preparation the particle size should be limited. The available literature data do not allow us to define an exact limit; however, the use of particles smaller than approximately 150 μm is an indication rather than a rule.

It is, of course, assumed that no agglomeration is taking place. The smaller the particles the less the possible mechanical irritation to the patient (esp. < 50 μm) and the higher the dissolution rate, and therefore drugs with a low water solubility will be dispensed in small, preferably micronized, particles. One should, however, be aware of the increased tendency of these particles to agglomerate as a result of strongly increased van der Waals forces in this case. Also, an unnecessary size reduction operation should be avoided if possible.

There are good indications that size reduction is not a good decision for all drugs. It has been shown, especially for readily water-soluble drugs, that large particles give blood levels that are higher than or at least equivalent to small particles. This would lead to the suggestion to use particles in the size range 50–100 μm. The lower limit of 50 μm to increase transport through the molten vehicle (see Fig. 34.2) and the upper limit of 100 μm is a safe protection against undue sedimentation during preparation. There is, however, no clear-cut picture, as to which solubility class this would apply to. For example, paracetamol (solubility in water approximately 15 mg mL^{-1}) gave the best blood levels when the particle size was smaller than 45 μm.

It should also be borne in mind that the spreading suppository mass should drag the suspended particles along to maximize the absorption surface. For heavy compounds it has become clear that this is a problem, but so far little or no proof is available that organic drugs (density usually 1.2–1.4 g cm^{-3}) suffer from this disadvantage when dispersed in, for example, 150 μm sized particles. Principally this may be expected, but care is needed to prevent too-rapid formulation decisions in this respect.

Amount of drug A complicating factor is the amount of drug present in a suppository. If the

number of particles increases, this would also increase the rate of agglomerates formation. This will depend very much on particle size and the presence of additives. The theory describing the agglomeration behaviour of dispersed systems (DLVO theory, see Chapter 6) can be applied in the non-aqueous systems we are dealing with, but certain refinements are necessary. Another consequence of the presence of suspended particles is the increased viscosity of the molten base. Also in this case we have to rely largely on empirical data, rather than on theory. It therefore seems advisable to include a decision on particle size in the development plan for an actual suppository formulation.

Other additives

For several widely varying reasons, formulators of suppositories make use of additives to improve their product. Most of these additions are based on empirical data and will be dealt with in the accompanying volume on dispensing (in preparation). The dispensing aspects include formulations for specific drugs that affect the melting point of the suppository; it may become depressed (by a soluble liquid compound) or increased (by a high amount of soluble high-melting active compound). The important point to consider in these situations is the possible influence of formulation changes on the release characteristics. We will further limit the discussion to fatty suppositories, where this plays a particular role.

The addition of viscosity-increasing additives (e.g. colloidal silicon oxide or aluminium monostearate, both approximately 1–2%) will create a gel-like system with a slower release rate of the drug. Data from the literature are not consistent on this point. In vitro this can be easily established, but whether the actual release in vivo will also be depressed cannot be easily predicted, as rectal motility will in certain cases be able to overcome this problem. The addition of lecithin is a worthy possibility when high amounts of solid drug are used. The reason why has clearly to be found in a decreased attraction between the drug particles, altering the flow properties of the dispersion in the positive sense.

The addition of surface-active agents has been extensively practised but still remains a source of great uncertainty. When these compounds are used to create an emulsion system (thus w/o) this must certainly be discouraged, as the release will be unacceptably slow. It may well be, however, that surfactants act as wetting agents. This can influence the release in a positive sense, but so far very little convincing information is available showing that wetting (i.e. displacement of base by rectal fluid) is a real problem. Surfactants may also act as 'deglomerators', which may prevent the formation of cake in the melting suppository, which in turn would certainly slow down drug release. Also here no firm conclusions can be drawn, as very little research work has been performed on agglomeration in non-aqueous media. The role of surfactants as spreading enhancers has never been clarified either, and this factor is strongly related to the occurrence of rectal motility. There are good indications that the presence of surfactants in a concentration higher than the critical micelle concentration can retard drug release from the suppository.

The finished product

Manufacture

Suppositories are manufactured both on a small scale in batches of 10–20 and on a (semi)automatic scale in batches up to 20 000 per hour. Essentially the mode of manufacture is similar in both cases, and involves melting of the vehicle, mixing the drug and the molten vehicle, dispensing in a former, cooling to solidify and, if necessary, packing in the final container. This includes a number of technological processes for which the relevant theory should be considered (see, for example, Chapter 13 for the mixing of semisolids). Most suppositories are nowadays packed individually in a plastic (PVC) or aluminium foil pack. Requirements leading to a good protection against moisture and oxygen can be deduced from the individual needs of the drug and the properties of the packaging material.

Quality control

A list of properties that should be controlled is given in Table 34.6.

The ***appearance*** of a suppository includes its odour, colour, surface condition and shape. These

Table 34.6 Control parameters of suppositories

Appearance
Weight
Disintegration
Melting (dissolution) behaviour
Mechanical strength
Content of active ingredient
Release

are organoleptically controlled and will be discussed in the accompanying volume (in preparation). The requirements for ***weight*** and ***disintegration*** are given in the European and national pharmacopoeias. The melting and dissolution behaviour is in fact reflected in the disintegration test. Many other methods are available, but none of them has been shown to provide more relevant information. The European Pharmacopoeia method proves to be rather insensitive and not too much value should be placed on the passing of this test.

The ***mechanical strength*** can be valuable to avoid problems with formulations in which the melting range has been depressed. This can be tested in several ways, including a tablet-crushing strength tester.

No official requirements are published as yet with regard to ***content uniformity***. This has, however, been shown to be a potential problem in the semi-automatic manufacture of batches of a few kilograms, when, owing in particular to sometimes insufficient mixing, the uniformity of content was insufficient. Paying careful attention to the design and control of the filling apparatus could solve most of these problems. However, the beginning and the end of the production process still give poor results, necessitating some degree of rejection.

Drug release from suppositories

Perhaps the most important thing to realize is that for the patient the release characteristics are the determining step towards the success of the therapy. What is really wanted is optimal bioavailability (for details see Part Three), which for the formulator means ensuring optimal and reproducible release in vivo.

Because there are very few ways to obtain in vivo release information this will usually have to be interpreted from in vitro release, which introduces the problem of in vitro/in vivo correlation. Current knowledge does not permit the choice of an in vitro method with a high predictive power for in vivo performance. Some aspects can be discussed, however, to give helpful pointers in this respect. Table 34.7 lists the parameters to be examined in testing suppository release in vitro.

Table 34.7 **In vitro release parameters**

Temperature
Contact area
Release medium
Movements
Membranes

The ***temperature*** to be chosen for testing rectal dosage forms is easily defined as the body temperature. Although for most practical purposes this can be set at 37°C, this is not the case for especially fatty suppository testing, for example. Most available vehicles have melting ranges below 37°C, but this does not necessarily mean that their viscosity at 37°C is the same. As the body temperature may be as low as 36°C at night, this implies that the release rate at 37°C may be an overestimate. Also, comparing bases at 37°C may lead to erroneous conclusions. The temperature at which testing is performed might be crucial, especially when ageing has occurred. Special attention should therefore be given to the actual testing temperature.

In the set-up shown in Figure 34.3 the temperature at the surface of the water layer inside the tube, where molten suppository material is gathered, may be a few degrees lower than the bulk temperature. By choosing the right dimension and closing the tube on the upper side this problem is eliminated here.

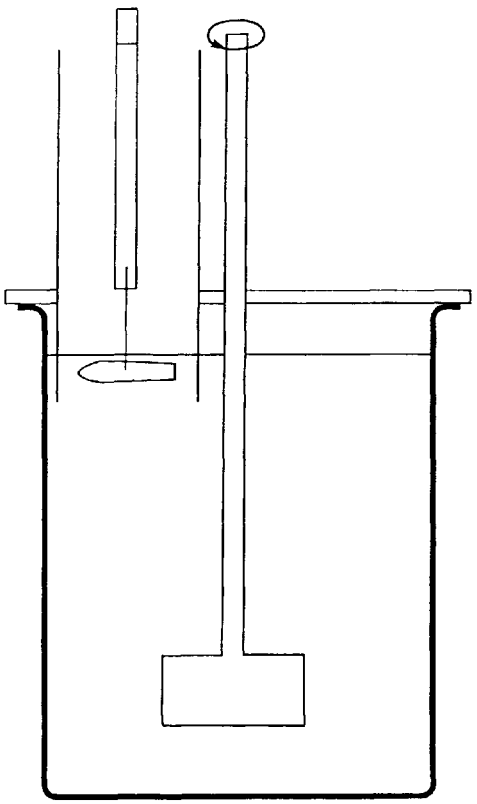


Fig. 34.3 Tube apparatus for sedimentation-controlled release testing from suppositories.

The ***contact area*** in the rectum over which spreading occurs cannot be standardized without introducing either an over- or an underestimate. In Figure 34.3 the area is relatively small (i.e. approximately 10 cm^2) compared to the total surface area of the rectum (approximately 300 cm^2). This type of apparatus therefore is clearly intended to be used for comparative studies only, and not for a complete in vivo simulation. At present no method is available that closely mimics the in vivo situation.

Another parameter to be considered is the ***release medium***. Because not enough information is available on the actual composition and structure of the rectal fluid, a choice is usually made for a relatively large volume of water or buffer solution. As the rate-limiting step in the bioavailability of fatty suppositories is very often drug release from the suppository, it seems reasonable enough not to include mucins that control the viscosity in vivo. The large volume in most in vitro methods would then not be so important either. More difficult is the choice of buffer, and especially its strength, as little is known about this factor in vivo (see above). For water-soluble vehicles the problem is even greater, and essentially no ideal solution has yet been found to the problem of choosing the volume and composition of the release medium. Interest has been created in the inclusion of one or another way of incorporating a ***pressure*** feature in release testing. It is clear that rectal motility exists and that it may influence bioavailability, but it is not yet clear how to incorporate this knowledge in the design of a release tester. Attempts have been made, but no conclusive answer has yet been found.

Very often ***membranes*** have been used in release testers, usually to envelop the suppository in a small volume of release medium. This has the enormous drawback that the release as measured in the outer compartment is not equal to the actual release taking place in the inner compartment. Most published results do not take into account that the membrane may form a resistance to passing drug molecules, and that the actual release may be underestimated. By a calculation procedure it is possible to obtain the actual release if certain conditions can be met. It seems advisable, therefore, to avoid membranes in a release tester whenever possible.

The actual true validation of in vitro release testing remains the in vivo performance. Several possibilities exist to obtain such data. Bioavailability determination should consider both rate and extent of absorption. Whenever possible these data should be obtained in humans, as at present no sufficiently reliable animal model is available. For a more detailed discussion on the general aspects of bioavailability testing and in vitro/in vivo correlations, see Chapter 18.

Rectal formulations other than suppositories

Apart from suppositories, many formulations can be used for the rectal administration of drugs. For the treatment of ***local*** disturbances, such as haemorrhoids, fatty ointments are widely used. In the treatment of rectocolitis large-volume enemas are used, e.g. 100 mL. This enables the drug to reach the upper part of the rectum and the sigmoid colon.

For the ***systemic*** administration of drugs, delivery forms such as tablets, capsules and microenemas are used. Tablets are not very attractive because they cannot disintegrate rapidly, owing to the small amount of water present in the rectum. Tablets that release CO_2 after insertion can be used, thereby stimulating defecation.

Capsules used to achieve a systemic effect are usually filled with a solution or suspension of the drug in vegetable oil or paraffin. Such capsules are mostly of the soft-shell type. Limited experience has been obtained with this dosage form, but it seems that there are no striking differences between the bioavailability from rectal capsules and that from fatty suppositories.

Microenemas are solutions or dispersions of the drug in a small volume (approximately 3 mL) of water or vegetable oil. This form is supplied in a small plastic container equipped with an application tube. After insertion of the tube, the container is emptied by compressing the bulb. The advantage of this delivery system is obvious, as no melting and dissolution process is necessary before drug release can start, if water is used as a vehicle. Many good results have been obtained with drugs delivered in microenemas, but this form is still of limited applicability because of its relatively high cost compared to suppositories, for example. Moreover, administration cannot be performed easily by patients themselves, and it is rather difficult to deliver the total content of the plastic container.

VAGINAL DRUG DELIVERY

Vaginal administration of drugs

The vaginal route is mainly used for the achievement of local effects, e.g. in the case of *Trichomonas* and *Candida* infections. Some drugs are, however,

administered vaginally to achieve systemic effects. In some cases the drugs given by intravaginal route have a higher bioavailability than with the oral route, because the drug enters immediately into the systemic circulation without passing the metabolizing liver (as is the case with drugs absorbed from the lower part of the rectum).

The vaginal wall is very well suited for the absorption of drugs for systemic use, as it contains a vast network of blood vessels. Only a few drugs are administered by this route at present, however. Among these are oestrogens and prostaglandin analogues, which are usually administered as vaginal creams or hydrogels. Progesterone has been given as vaginal suppositories (pessaries) for some years, and better results are obtained in this way than after oral dosing.

Formulation of vaginal dosage forms

Many different types of formulations have been and are applied vaginally, e.g. tablets, capsules, pessaries, solutions, sprays, foams, creams and ointments. Because of the rather low moisture content under normal physiological conditions, additives are used to improve the disintegration of vaginal tablets, e.g. bicarbonate together with an organic acid, which results in CO_2 release. A good filler for vaginal tablets is lactose, as this is a natural substrate for the vaginal microflora, which converts lactose into lactic acid, resulting in a pH value of 4–4.5.

Vaginal suppositories (pessaries) are mostly prepared with glycerol–gelatin bases, as this mixture is well tolerated. Polyethylene glycols are less common because they are said to promote irritation. Also, fatty excipients are not much used. Most delivery forms for vaginal application demand an auxiliary device to obtain deep insertion of the delivery system.

BIBLIOGRAPHY

de Blaey, C.J. and Polderman, J. (1989) Rationales in the design of rectal and vaginal delivery forms of drugs. In: *Drug Design*, Vol. 9 (ed. E.J. Ariëns). Academic Press, New York, pp. 237–266.

de Boer, A.G., Molenaar, F., de Leede, L.G.J., Breimer, D.D. (1982) Rectal drug administration: clinical pharmacokinetic considerations. *Clin. Pharmacokinetics*, 7, 285–311.

Müller, B. (ed) (1986) *Suppositorien*. Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart.

Rytting, J.H. (1990) Rectal route of peptide and protein delivery. In: *Peptide and Protein Drug Delivery* (ed. V.H. Lee). Marcel Dekker, New York, pp 579–594.

Thoma, K. (1980) *Arzneiformen zur rectalen und vaginalen Applikation*. Werbe- und Vertriebsgesellschaft Deutscher Apotheker mbH, Frankfurt am Main, Germany.

Tondury G. (1981) *Angewandte und Topographische Anatomie*, 5th edn. Georg Thieme Verlag, Stuttgart.

Winfield, A.J. and Richards, R.M.E. (1998). *Pharmaceutical Practice*, 2nd ed. Churchill Livingstone, Edinburgh.

35
Delivery of pharmaceutical proteins

Daan Crommelin, Ewoud van Winden, Albert Mekking

CHAPTER CONTENTS

INTRODUCTION

Protein structures

Pharmaceutical proteins are built up of amino acids chains (their primary structure). To be pharmacologically active, this amino acid sequence must form a well defined three-dimensional structure. Parts of the protein will fold in locally identifiable, discrete structures such as α helices or β sheets (known as secondary structures). The overall (tertiary) structure of the protein is established by the proper positioning of the different subunits relative to each other. In some cases individual protein molecules form a quaternary structure, in which the individual protein molecules interact and build a larger, well defined structure (e.g. haemoglobin).

The formation and stability of the secondary, tertiary and quaternary structures is based on relatively weak physical interactions (e.g. electrostatic interactions, hydrogen bonding, van der Waals forces and hydrophobic interactions) and not on covalent chemical-binding principles. Repulsive energy between apolar parts of the protein and water are responsible for hydrophobic interactions. The physical forces involved are relatively weak. This means that protein structures can be rather easily changed, leading to a modification or even loss of their pharmacological characteristics.

Amino acid chains can be modified by covalently attaching non-amino acid sections, such as sugar (glycoproteins), phosphate or sulphate groups. In particular, the sugar part can make up a substantial part of the molecular weight of the (glyco)protein. These groups may be essential for the pharmacological effect of a therapeutic protein, not only while acting at its receptor sites, but also to provide the proper pharmacokinetic profile.

Pharmaceutical protein molecules are large and diffusional transport through epithelial barriers such

as those encountered in the gastrointestinal tract is slow unless specific transporter molecules are available. Moreover, the conditions in the lumen of the GI tract are extremely hostile to these proteins. (Enzymatic) degradation is fast. Therefore, the large majority of pharmaceutical proteins are delivered via the parenteral route (i.e. by the needle). The issue of alternative routes of administration is discussed later in this chapter.

Conserving the integrity of these large molecules is essential to ensure an optimal therapeutic effect and to minimize effects such as the induction of unwanted immune responses. An immune response may neutralize the therapeutic activity in chronic dosing schedules and cause serious side-effects. Protein stability concerns both their chemical and their physical structure. There are many functional groups in the amino acid chain available for chemical degradation, and the preferred three-dimensional structure is readily irreversibly disturbed (e.g. through heat, changes in pH or ionic strength). Some analytical approaches to monitor the protein structure are discussed later.

The preferred shelf-life for pharmaceutical products is a minimum of 2 years. Most proteins degrade too fast when formulated as aqueous solutions, even when kept in the refrigerator. Therefore, they have to be stored in a dry form and be reconstituted before administration. These delicate structures are usually dried by freeze-drying (see Chapter 26). The choice of the proper excipients (e.g. lyoprotectants) has proved to be extremely important.

Sources of pharmaceutical proteins

Nowadays most proteins used in therapy or under development are produced by recombinant DNA or hybridoma technology (known as biotechnology or biotech products). Examples are human insulin, erythropoietin, monoclonal antibodies, cytokines and interferons. They are all produced in cell cultures by prokaryotic or eukaryotic cells, ranging from *Escherichia coli* to mammalian cells such as Chinese hamster ovary cells, or transgenic animals. From the examples listed, one may conclude that many of the pharmaceutical proteins are basically endogenous products. However, a number of currently used biotech products are not exactly identical to the endogenous product. For example, the glycosylation patterns of the recombinant form may be reproducibly produced on a large scale, but not completely match the endogenous product. Extensive evaluation of these products in clinical trials has proved their efficacy and safety.

Isolation of the expressed protein from the culture medium is a multistep process consisting of several different (chromatographic/filtration) steps. For every protein a 'tailor-made' purification protocol has to be developed to remove impurities while ensuring integrity.

Biotech-derived molecules may make up the majority of protein drugs, but there are still proteins of major therapeutic importance isolated from blood from humans or animals. Examples are albumin, blood clotting factors (such as Factor VIII from blood from human volunteers), and antisera from patients or animals such as horses and sheep. Here again, special purification protocols have to be developed, with particular emphasis on reduction of viral contamination (see later).

Specific challenges

It is clear that pharmaceutical proteins offer special challenges to the pharmaceutical formulator. They are delicate, large molecules with many functional groups. Their structure, being stabilized by relatively weak physical bonds, is readily and irreversibly changed. In vivo this may directly affect the interaction with the receptor, change their pharmacokinetic characteristics, e.g. their clearance, or make them immunogenic. Moreover, their epithelial penetration capability is very low unless the proper transporter molecules are available. Thus, as a rule, pharmaceutical proteins are administered parenterally.

In the sections that follow several issues will be dealt with in more detail.

FORMULATION OF PHARMACEUTICAL PROTEINS FOR PARENTERAL ADMINISTRATION

Stability issues

In the introduction to this chapter it was pointed out that pharmaceutical proteins are high molecular weight molecules with amino acid building blocks that are sensitive to degradation and with a specific three-dimensional structure. Table 35.1 lists the pathways for degradation of proteins.

Physical instability

Degradation rates depend on environmental conditions and the formulator should carefully select conditions for optimal stability. For example,

Table 35.1 Pathways of degradation of proteins.

Fragmentation

H_2N — Lys — Ala — COOH → H_2N — Lys + Ala — COOH

Isomerization

L-form amino acid → D-form amino acid

Deamidation

H_2N ..... Asn ..... COOH → H_2N ..... Asp ..... COOH + $—NH_2$

Oxidation

H_2N ..... Met ..... COOH → H_2N ..... Met(=O) ..... COOH

Disulphide scrambling

H_2N ..... Cys ..... Cys ..... Cys ..... COOH ⟷ H_2N ..... Cys ..... Cys ..... Cys ..... COOH

H_2N ..... Cys ..... Cys ..... Cys ..... COOH

Oligomerization

n(H_2N ..... Cys ..... Cys ..... COOH) → n(H_2N ..... Cys ..... Cys ..... COOH)

n(H_2N ..... Cys ..... Cys ..... COOH)

Aggregation

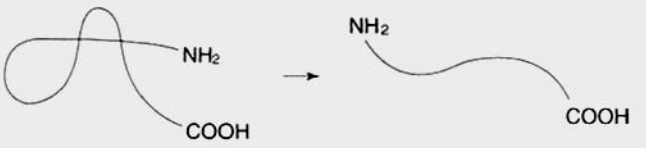
n(H_2N ..... COOH) → (H_2N ..... COOH)$_n$

Cross-linking

n(H_2N ..... COOH) → H_2N ..... COOH

H_2N ..... COOH

Denaturation



elevated temperatures can cause denaturation of proteins in aqueous solution. Interestingly, low temperatures may also induce destabilization. Besides, protein aggregation is often initiated by adsorption of the protein monomer on the walls of the container. Proteins may also aggregate by shaking or by exposure to shear forces. Hydrophobic parts of the molecule are then exposed to hydrophobic interfaces (air/water), the protein unfolds and aggregation occurs.

Chemical instability

Because of the many amino acids involved, full prevention of all chemical degradation reactions is difficult. The formulator should consider which chemical degradation pathways are relevant. Under neutral conditions the peptide bonds between amino acids are stable; only the asparagine–glycine and asparagine–proline bonds are relatively labile.

Deamidation is a rather common degradation reaction in water. Asparagine and glutamine are the amino acids that can be deamidated. Deamidation reaction kinetics depend on pH and neighbouring amino acids.

Oxidation is not limited to methionine and cysteine (Table 35.1): histidine, tryptophan and tyrosine are also sensitive to oxidation reactions. Oxidation is catalysed by traces of transition metal ions. An oxidative milieu may also cause free cysteine units to form disulphide bridges or disulphide bond scrambling.

Naturally occurring amino acids are in the L form. Isomerization to the D form is possible and will change the structure of the protein.

Improper choice of excipients may also cause degradation reactions. For example, sugars are often used as excipients (Table 35.2), but reducing sugars can react with free primary amino groups of the protein molecule via the so-called Maillard reaction (even in the dry state) and form brownish reaction products. Reducing sugars (e.g. lactose) should therefore be excluded from protein-containing formulations.

Excipients used

Table 35.2 lists the excipients used in protein-containing parenteral dosage forms. Not all of the ingredients listed are always needed, e.g. many pharmaceutical proteins are sufficiently soluble in water. This is in particular true for highly glycosylated molecules. In this case no solubility-enhancing substances are needed. However, if solubility enhancement is necessary, the selection of the proper pH conditions should first be considered. Protein solubility depends on its net charge. In general, as with low molecular weight drugs, uncharged protein molecules (at the pH of their isoelectric point, i.e.p.) have the lowest solubility in water. Therefore, choosing pH conditions 'away' from the i.e.p. can solve the protein (and the problem). Some amino acids (e.g. arginine and lysine) increase protein solubility and reduce aggregation reactions by a not-well understood mechanism. Detergents such as polysorbate 20 and 80 or sodium dodecyl sulphate can also be used to prevent aggregation. These compounds prevent the adsorption of proteins to interfaces (air/water and container/water) and thereby interface-induced protein unfolding. Human serum albumin has a strong tendency to adsorb to interfaces and may therefore be added to therapeutic protein formulations as an anti-aggregation agent.

Table 35.2 Excipients used in parenteral dosage forms and their function

Excipient	Function	Examples
Solubility-enhancing substances	Increase solubility of proteins	Amino acids, detergents
Antiadsorbent/aggregation blockers	Reduction of adsorption and aggregation prevention	Albumin, detergents
Buffer components	Stabilizing pH	Phosphate, citrate
Preservatives	Growth inhibition in vials for multiple dosing	Phenol, benzylalcohol, organic Hg-compounds
Antioxidants	Prevent oxidation	Ascorbic acid, sulphites, cysteine
Stabilizers during storage (lyoprotectants)	Preservation of integrity while in dry form	Sugars
Osmotic compounds	Ensure isotonicity	Sugars, NaCl

Oxidation reactions are catalysed by heavy metals. Chelating agents are used to reduce oxidation damage through binding of the ions. This approach cannot be used if the metal ion is necessary as an integral part of the protein structure. Examples are zinc ions in insulin formulations and iron ions in haemoglobin. Then, antioxidants such as sulphites may be added to reduce the oxidation tendency. In the case of vials for multiple dosing, preservatives have to be included in the formulation. Benzyl alcohol and phenol are often used for this purpose.

Buffered aqueous protein solutions may be stable for 2 years under refrigerator conditions. Some monoclonal antibody formulations, for example, are available as aqueous solutions, but the more common situation is that the formulation has to be freeze-dried in the vials to avoid degradation and to ensure that the product can be readily reconstituted.

During freeze-drying (Chapter 26) water is removed by sublimation. In the freeze-drying process three discrete phases can be discerned. The first is freezing of the solution to temperatures typically around −35 to −40°C, followed by a sublimation phase with temperatures of around −35°C and low pressures to remove the frozen water (phase 2), and a final, secondary drying stage to remove most residual water. The pressure must remain low, but the temperature can rise up to about 20°C without collapse of the porous cake (see below). A lyoprotectant (e.g. sugar) is necessary to stabilize the product as the removal of water may irreversibly affect the protein structure. Moreover, sugar lyoprotectants also happen to form readily reconstitutable porous cakes.

The freezing temperature should be low enough to convert the aqueous solution with the sugar and the protein into a glass. Glass formation in sugar solutions usually occurs around −30°C. Just below the glass transition temperature the sublimation process can begin during the lowering of the pressure in the chamber. The sublimated water is collected on a condenser with a considerably lower temperature (typically −60°C). As sublimation extracts a large amount of latent heat from the system, the temperature in the vials containing the frozen protein solution could fall even lower than the starting temperature, slowing down sublimation. The vials are therefore heated in a controlled way to keep them at temperatures low enough to preserve their glassy texture, but high enough to let the sublimation process proceed at a sufficiently high speed.

The mechanism(s) of action of lyoprotectants (non-reducing sugars) are not fully understood. The following may play a role:

- Lyoprotectants replace water as stabilizing agent ('water replacement theory') of the protein;
- Lyoprotectants increase the glass transition temperature in the frozen system and in the dried system, avoiding collapse of the porous cake which would slow down water removal from the frozen cake (during freeze drying) and interfere with a rapid reconstitution of the freeze-dried cake;
- Lyoprotectants slow down the secondary drying process and minimize the chances of overdrying of the product in the secondary drying stage.

Microbiological requirements

Typically, pharmaceutical proteins are administered via the parenteral route. This implies that the product should be sterile. In addition, virus and pyrogen removal steps should be part of the purification and production protocol.

Pharmaceutical proteins cannot be sterilized by autoclaving, gas sterilization or ionizing radiation, because these procedures damage the molecules. Therefore, sterilization of the end-product is not possible. This leaves aseptic manufacturing as the only option.

All utensils and components must be presterilized (by heat sterilization, ionizing radiation or membrane filtration) before assembling the final formulation to minimize the bioburden. Protein products are manufactured under aseptic conditions in class 100 areas (fewer than 100 particles > 0.5 μm per cubic foot). This low level contamination is reached by filtration of air through HEPA (high-efficiency particulate air) filters. Finally, the product is filled into the containers through sterile filters with 0.22 μm pores before capping or freeze drying/capping.

Pharmaceutical proteins are produced by living organisms. Viruses can be introduced into the product either by the use of contaminated culture media or via infected (mammalian) production cells. It is therefore important that purification and manufacturing protocols contain viral decontamination steps. Viral decontamination can be accomplished by virus removal and/or by viral inactivation. The problem faced when selecting inactivation techniques is that there is often a narrow window between successful viral inactivation and preservation of the integrity of the pharmaceutical protein structure.

Viruses can be removed by filtration, precipitation or chromatography. For virus inactivation, heat treatment (pasteurization), radiation or crosslinking agents (e.g. β-propiolactone) can be used. As no single process guarantees complete virus removal, often

several different decontamination steps are introduced in series in the 'downstream' purification process and in the manufacturing of the final formulation.

Gram-negative host cells, such as *E. coli*, are often used as production cells for non-glycosylated proteins. Gram-negative cells contain large amounts of endotoxins in their membranes. These endotoxins are heat stable, amphipatic, negatively charged lipopolysaccharides and are potent pyrogens. Pyrogens have to be removed in order to meet pharmacopoeial criteria, and this can be done, for example, through anion-exchange chromatography.

ANALYTICAL TECHNIQUES TO CHARACTERIZE PROTEINS

It is clearly important to be able to guarantee the integrity of a protein. As mentioned earlier, a protein molecule is a complex three-dimensional structure of amino acids, often coupled to saccharide, phosphate or sulphate moieties. The total structure is responsible for the pharmacodynamic (e.g. receptor interaction) and pharmacokinetic (e.g. clearance, targeting) effect. It is not possible to define the structure of a pharmaceutical protein with the same precision as small, low molecular weight molecules, where a combination of analytical techniques provides unequivocal structural evidence.

Therefore, a set of pharmacological, immunological, spectroscopic, electrophoretic and chromatographic approaches is used to characterize the protein as closely as possible. Table 35.3 lists a number of regularly used analytical techniques and the information that is obtained.

Quality assessment used to be based on functional tests in vivo (relevant animal models). An example is the pharmacopoeial test for insulin: the lowering of the blood glucose level in rabbits upon

Table 35.3 Approaches to confirm protein structure

Approach	Information obtained
In vivo tests, use of test animals	Pharmacological effect
In vitro tests (sensitive cells)	Functional test
Immunological tests	
ELISA	Interaction with one epitope on protein
RIA	Interaction with one epitope on protein
Analytical approaches	
Spectroscopic	
UV spectroscopy	Secondary/tertiary structure
fluorimetry	Secondary/tertiary structure
CD spectroscopy	Secondary/tertiary structure
infrared spectroscopy	Secondary/tertiary structure
mass spectrometry	Secondary/tertiary structure
Electrophoretic approaches	
SDS-PAGE	Molecular weight
IEF	Isoelectric point
High-performance liquid chromatography (HPLC)	
GP (gel permeation)	Molecular weight/aggregates
HI (hydrophobic interaction)	Hydrophobic interactions
Affinity chromatography	Interaction with specific ligand
IEC (ion exchange)	Charge patterns
RP (reversed phase)	

CD, circular dichroism;
ELISA, enzyme-linked immunosorbent assay;
IEC, ion-exchange chromatography;
IEF, isoelectric focusing;
HPLC, high-performance liquid chromatography;
MALDI, matrix-assisted laser desorption ionization;
MS, mass spectrometry;
RIA, radioimmunoassay;
SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis;

injection of the insulin product to be tested. These tests do not have the sensitivity to identify small changes in molecular structure or detect early degradation products, and they do not provide information on such things as the presence of product immunogenicity. In vitro cell tests, such as those used for cytokine activity assessment, inform us about the functional activity of the molecule, but not its pharmacokinetic behaviour or immunogenicity. ELISA (enzyme-linked immunosorbent assay) and RIA (radioimmunoassay) belong to the class of immunological tests. Here the interaction of a monoclonal antibody with one epitope region on the protein is determined. The rest of the molecule is not 'probed'.

Electrophoretic techniques such as SDS-PAGE (sodium dodecyl sulphate–polyacrylamide gel electrophoresis) and IEF (isoelectric focusing) are powerful tools to assess product purity and to provide molecular weight and isoelectric point (i.e.p.) information regarding the protein.

Table 35.3 lists a number of chromatographic techniques that elucidate product characteristics. In particular, impurities and degradation products can be picked up at an early stage. Gel chromatography discriminates mainly on the basis of molecular size and is a powerful technique to monitor aggregate formation. Ion-exchange resins separate on the basis of subtle variations in protein charge patterns and are being used to detect oxidation (e.g. methionine) and deamidated (converted glutamine and asparagine) products. Modern mass spectroscopic techniques such as MALDI–TOF (matrix-assisted laser desorption ionization–time of flight) mass spectroscopic analysis, or a combination of HPLC (high-pressure liquid chromatography) with electrospray ionization-induced mass spectrometry give detailed information on amino acid sequence and glycosylation patterns.

In conclusion, to ensure pharmaceutical protein quality one must follow a strict protocol regarding the definition of the protein production cell lines used, the chosen culturing conditions and downstream processing conditions, and the filling/(drying)/ finishing process. Analytical approaches to confirm the protein structure will always include a long list of approaches, ranging from in vivo tests in animals to information provided by highly sophisticated analytical technologies.

None of these tests tells the whole story; together they tell more, but there is never the situation encountered with many low molecular weight molecules whereby a full description of the drug, including a detailed impurity profile, is available.

ADMINISTRATION OF PHARMACEUTICAL PROTEINS

Routes of administration

As mentioned in the introduction to this chapter, oral administration of a pharmaceutical protein results in a very low bioavailability. The protein is enzymatically attacked in the gastrointestinal tract and, moreover, penetration through the gut wall will be slow and incomplete. Oral vaccines containing antigenic protein material are an exception to the general rule that proteins should not be administered orally. With vaccines, even low uptake levels may still deliver sufficient material to lymphoid tissue just below the epithelium (in the so-called Peyer's patches) to induce a strong (both local and systemic) immune response.

When a protein is delivered intravenously clearance from the blood compartment can be fast, with a half-life of minutes, or slow, with a half-life of several days. An example of a rapidly cleared protein is tissue plasminogen activiator (tPA), with a plasma half-life of a few minutes. On the other hand, human monoclonal antibodies have half-lives of the order of days.

Protein drugs are often administered subcutaneously or intramuscularly. These routes of administration are considered to be more patient friendly and the injection process easier than with the intravenous route. Upon intramuscular (i.m.) or subcutaneous (s.c.) injection the protein is not instantaneously drained to the blood compartment. Studies monitoring the fate of a protein upon s.c. injection demonstrate that passage of a protein through the endothelial barrier lining the local capillaries at the site

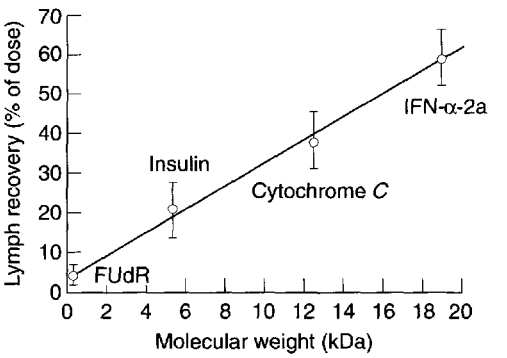


Fig. 35.1 Correlation between the molecular weight and the cumulative recovery of recombinant interferon (IFN α-2a), cytochrome C, inulin and 5-Fluoro-2'-deoxyuridine (FudR) in the efferent lymph from the right popliteal lymph node following s.c. administration into the lower part of the right hind leg of sheep (from Crommelin and Sindelar, 1998)

of injection is size dependent. If the protein is too large it will enter the lymphatic system and be transported via the lymph into the blood. Figure 35.1 shows the relationship between molecular size and lymphatic drainage. Lymphatic drainage takes time and a delay in the onset of systemic activity is observed. The protein is also exposed to the local environment containing proteases. Therefore, the bioavailability of protein drugs upon s.c. (and i.m.) administration can be far from 100%. This can have dramatic consequences, e.g. some diabetics become insulin resistant because of high tissue peptidase activity.

There is not always a direct relationship between plasma level and pharmacological response (i.e. no direct pharmacokinetic–pharmacodynamic (PK/PD) relationship). As the mechanism of action of a drug might be complex, involving different sequential steps, fast clearance from the blood compartment may not necessarily mean that drug action is also short-lived. The relationship between a pharmacokinetic profile and the pharmacodynamic result of the presence of the drug can be quite complex. A drug may trigger a reaction, which may result in measurable, pharmacological effects much later. As an example, the cytokine intraleukin-2 (IL-2) (in its PEG-ylated form) is rapidly cleared from the blood compartment and a pharmacological effect (increase in the number of blood lymphocytes) is observed long afterwards (can be days) (Fig. 35.2).

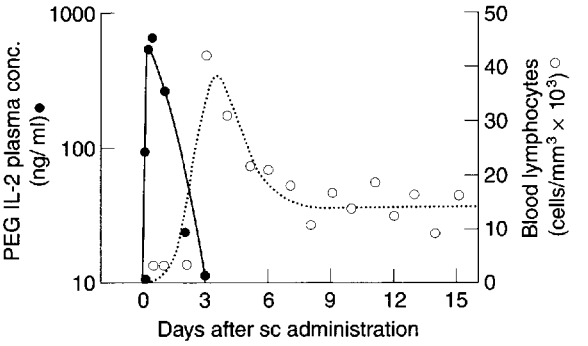


Fig. 35.2 PEG-IL-2 pharmacokinetics and pharmacodynamics (changes in blood lymphocyte count) after subcutaneous adminstration of 10 MIU/kg in rats. PEG = poly(ethylene glycol); (from Crommelin and Sindelar 1998)

Finding alternatives for the parenteral route has been an area of interest for many years. Table 35.4 lists different possible routes of delivery for proteins.

With the exception of the pulmonary route, all other options have a low bioavailability. Some bioavailability data on the intratracheal administration of proteins in rats are shown in Table 35.5. The extent of absorption depends strongly on the nature

Table 35.4 Alternative routes of administration to the oral route for biopharmaceuticals (adapted from Crommelin and Sindelar 1998)

Route	Relative advantage	Relative disadvantage
Nasal	Easily accessible, fast uptake, proven track record with a number of 'conventional' drugs, probably lower proteolytic activity than in the GI tract, avoidance of first-pass effect, spatial containment of absorption enhancers is possible	Reproducibility (in particular under pathological conditions), safety (e.g. ciliary movement), low bioavailability for proteins
Pulmonary	Relatively easy to access, fast uptake, proven track record with 'conventional' drugs, substantial fractions of insulin are absorbed, lower proteolytic activity than in the GI tract, avoidance of hepatic first-pass effect, spatial containment of absorption enhancers (?)	Reproducibility (in particular under pathological conditions, smokers/non-smokers), safety (e.g. immunogenicity), presence of macrophages in the lung with high affinity for particulates
Rectal	Easily accessible, partial avoidance of hepatic first-pass effect, probably lower proteolytic activity than in the upper parts of the GI tract, spatial containment of absorption enhancers is possible, proven track record with a number of 'conventional' drugs	Low bioavailability for proteins
Buccal	Easily accessible, avoidance of hepatic first-pass effect, probably lower proteolytic activity than in the lower parts of the GI tract, spatial containment of absorption enhancers is possible, option to remove formulation if necessary	Low bioavailability of proteins, no proven track record yet (?)
Transdermal	Easily accessible, avoidance of hepatic first-pass effect, removal of formulation is possible if necessary, spatial containment of absorption enhancers is possible, proven track record with 'conventional' drugs, sustain/controlled release possible	Low bioavailability of proteins

Table 35.5 Absolute bioavailability of a number of proteins (intratracheal vs intravenous) in rats

Molecule	MW (kDa)	No. of amino acids	Absolute bioavailability (%)
α-Interferon	20	165	> 56
PTH-84	9	84	> 20
PTH-34	4.2	34	40
Calcitonin (human)	3.4	32	17
Calcitonin (salmon)	3.4	32	17
Glucagon	3.4	29	< 1
Somatostatin	3.1	28	< 1

PTH, recombinant human parathyroid hormone.
(From Crommelin and Sindelar 1998 with permission)

of the protein. Insulin is a candidate drug for pulmonary delivery to diabetics to mimic the natural physiological response to a meal (postprandial glucose control). Subcutaneous injection gives a relatively slow response; pulmonary uptake is faster. New pulmonary delivery devices (see Chapter 31) not only increase average bioavailability but also reduce variation in uptake.

Three approaches have been followed to improve the bioavailability of pharmaceutical proteins when exploring alternative routes of administration. First, coadministration of protease inhibitors, such as bacitracin, should slow down metabolic degradation. Second, excipients (often with an amphipatic character, such as bile salts) can be added to enhance passage through epithelial barriers. The third approach is to prolong the presence of the protein at the absorption surface, e.g. by the use of mucoadhesives.

Intranasal delivery of chitosan and starch microspheres demonstrated enhanced uptake of coadministered insulin. In humans, intranasal delivery of insulin with chitosan results in absolute bioavailabilities of 7%. In conclusion, bioavailability is indeed strongly enhanced when using these approaches, but safety issues must be addressed before these absorption enhancers can be introduced into marketed products.

Release control

Many therapeutic proteins are short-lived in the blood compartment. Assuming there is a direct relationship between blood level and therapeutic effect, it is important to maintain therapeutically relevant drug concentrations in the bloodstream. Portable pump systems with adjustable pump rates are available for patients. Catheters provide the link between the pump and, for example, the peritoneal cavity. These systems are particularly useful if a constant dose input is required and the drug is needed over a limited period of time. Otherwise, more flexible delivery systems are preferred. For insulin (with a plasma half-life of 5 minutes) different forms of controlled-release systems for s.c. injection are available. Release control is based on different physicochemical appearances (amorphous/crystalline) of insulin itself and on insulin complexes with Zn^{2+} ions or proteins, such as protamine. Zn^{2+} ions tend to slow down the release of insulin. Amorphous insulin plus Zn^{2+} ions results in moderate prolongation of drug action. Crystalline insulin plus Zn^{2+} ions gives a long-acting product. The addition of protamine (at neutral pH a positively charged protein) to insulin and Zn^{2+} ion combinations protracts the insulin effects even more (up to 72 hours; long-acting). Isophane insulin (NPH: neutral protamine Hagedorn) contains insulin and protamine in isophane proportions (no excess of either component), resulting in intermediate-acting formulations.

At present, efforts are being made to build 'closed-loop' systems where insulin administration is controlled by:

- a biosensor permanently monitoring blood glucose levels;
- an infusion pump with adjustable pump rate; and
- an electronic section with an algorithm linking blood glucose levels to insulin need at any time.

In hospitals such equipment is available to stabilize blood glucose in patients for limited periods, but no portable 'patient-friendly' systems for chronic use are yet available.

In even earlier stages of development are 'artificial pancreases'. Isolated insulin-producing β cells from the islets of Langerhans are introduced into the body in a container system with a wall that allows the passage of glucose, insulin and nutrients. However, the container wall keeps the β cells separated from the patient's immune system. Increasing blood glucose levels will stimulate the secretion of insulin by the encapsulated β cells. The excreted insulin will be released from the container and glucose levels will fall until normal blood levels are reached.

Controlled-release systems containing microspheres (with diameters between 10 and 100 μm) for s.c. administration are currently under development. A 30-day action sustained-release system designed on the basis of biodegradable polylactic-glycolic acid

has been formulated for human growth hormone. These microspheres are prepared using a double emulsion technique whereby the proteins are exposed to organic solvents, and protein encapsulation efficiency is rather low. Alternatively, a dextran-based microsphere preparation protocol has been developed without the use of organic solvents and with extremely high loading efficiencies.

CONCLUDING REMARKS

Protein drugs are rapidly gaining in importance, and both market volume and market share are expected to rise. Biotechnological techniques permit the design and synthesis of active proteins. It is the task of the pharmaceutical formulation scientist to turn the pure substance into a formulation that can be safely administered to the patient, exerting optimal therapeutic benefits. In this chapter different aspects of this formulation process are described and special attention is given to those aspects where biotech products clearly differ from low molecular weight drugs.

BIBLIOGRAPHY

Baudys, M. and Kim, S.W., eds. (1999) Insulin Delivery. *Adv. Drug Del. Rev.* **35**, 141–335.

Crommelin, D.J.A. and Sindelar R.D. (1998) *Pharmaceutical Biotechnology*. Harwood Academic Publishers, Amsterdam.

Franssen, O., Vandervennet, P. Roders, P. and Hennink, W.E. (1999) Chemically degrading dextran hydrogels: controlled release of a model protein from cylinders and microspheres. *J. Controlled Release*, **59**, 219–228.

Manning, M.C., Patel, K. and Borchardt, R.T. (1989) Stability of proteins, *Pharm. Res.* **6**, 903–917.

Wang, W. (1999) Instability, stabilization and formulation of liquid protein pharmaceuticals. *Int. J. Pharm.* **185**, 129–188.

Van de Weert, M., Hennink, W.E. and Jiskoot, W. (2000) Protein instability in Poly(lactic-co-glycolic acid microparticles. *Pharm. Res.*, **17**, 1159–1167.

36
Packs and packaging

Dixie Dean

CHAPTER CONTENTS

Packaging can be defined as an economical means of providing presentation, protection, identification/information, containment, convenience and compliance for a product during storage, carriage, display and use until such time as the product is used or administered. This total timescale must be within the shelf-life of the product, which is controlled by the selection of the right combination of product and pack. In looking at the above definition emphasis on any one factor may change with time, advancements in science and technology, or trends in product form. For instance, there has been a distinct move from the use of unpleasant oral liquids to the solid dose form, which has recently concentrated on delayed- or sustained-release products. Such changes can have a positive influence on the type of pack used, as shown by the increasing applications of blister and strip packaging. Both sustained-release and these unit-type packs offer obvious patient convenience, as a selected number of units can be readily detached and carried as a day's treatment, possibly leading to improved compliance. However, packaging can offer convenience factors anywhere along its lifecycle, e.g. reel-fed materials used for the production of blisters and strips need relatively little storage space compared to any preformed bottle, which literally wastes space by storing air prior to being filled. Further examples of packs offering patient convenience include a range of unit dose presentations that permit immediate disposal after use, metered dose aerosols and nasal sprays which combine convenience with dosage control, squeeze eyedrop packs (compared to the earlier bottle, dropper and teat assembly) etc.

The role of the pack

The role of the pack and the packaging operation needs emphasis as the shelf-life of all pharmaceutical products, irrespective of whether they are ethicals, semiethicals or proprietaries (over-the-counter or OTC), is largely dependent on certain functions of the pack. The pack must be economical and therefore contribute to overall profitability; it must provide protection against climatic, biological, physical and chemical hazards; it must provide an acceptable presentation which will contribute to or enhance product confidence while at the same time maintaining adequate identification and information; and last but not least it must contribute in terms of convenience and compliance. Each of these aspects has to be considered against the total (shelf) life of the product, which involves periods of storage (static), carriage (motion), possible display and finally use or administration, directly by a patient or indirectly by a health professional. In certain cases the pack may form part of the administration system, as is seen with aerosols metered dose nasal pumps, prefilled syringes etc. As these total pack requirements become increasingly extensive and sometimes conflicting, it is not surprising to find that the final choice is inevitably a compromise, hence the question 'What is an ideal pack?' rarely has a simple answer. For example, the external image of the pack must not only compliment product confidence, provide clear and concise product identification, adequate information related to the contents (including legal requirements), the route of administration, storage conditions, batch number, expiry date, manufacturer's name and address and product licence number, but also assist in patient compliance. Producing an aesthetically acceptable design is therefore only one stage where compromise has to be exercised.

In addition to this aesthetic requirement the majority of the remaining pack factors are associated with its function. The primary pack consists of those packaging components that form the part of the pack directly containing the product (i.e. bottle, cap, cap liner, label etc.). The main functions of the primary pack are to contain and to restrict any chemical, climatic or biological or occasionally mechanical hazards that may cause or lead to product deterioration. Because the primary pack also represents the pack of 'use', it must also function in the hands of the user as a means of drug administration. The packaging external to the primary pack, known as the secondary packaging, mainly provides the additional physical protection necessary to ensure the safe warehousing and delivery of the product to the point where bulk quantities are broken down into individual or specific units.

THE PACK AS A PROTECTION

Although each function of the pack has to assume a certain level of importance, protection is almost invariably the most critical factor as it controls the total shelf-life of the product. Asking the question 'protection against what?' produces a whole list of possible hazards, many of which are listed below. These are not identified in any particular order of importance as those that are relevant to a specific product will vary in both number and criticality. They cover mechanical, climatic, biological and chemical factors.

Mechanical hazards

Physical or mechanical damage may occur due to the following.

Shock or impact damage

This phrase implies rough handling, where rapid deceleration occurs (drops, impacts). Shock can normally be reduced or overcome by various forms of cushioning, restriction of movement, more careful handling etc. However, it should be noted that damage can occur to the pack or packaging material before it reaches the stage of a packed product.

Compression

Top pressure or loading can distort and crush a pack and damage the product inside. The crushing of a carton can make a product unsaleable even though no damage has occurred to the contents. Although this is most likely to occur during stacking in the warehouse or in transit, where vibration adds a further hazard, compression of the pack can occur in other situations (i.e. capping on a production line, when being carried home by the user etc.).

Vibration

Vibration consists of two variables, frequency and amplitude. These can vary enormously, e.g. a load on a truck may bounce up and down say 0–50 mm, up to 120 times per minute, whereas vibration from aircraft/ships engines may have an extremely low amplitude but a very high frequency. Each extreme may produce different forms of damage to product and/or pack, i.e. components of product may separate, screw caps may loosen, labels or decoration may abrade etc.

Abrasion

Although this results from both regular and irregular forms of vibration it is listed separately as the visual appearance of the product or pack can be affected, e.g. a rectangular bottle in a carton will move up and down and from side to side. A round bottle in the same circumstances will suffer from an additional possibility of rotation. See below under Chemical hazards.

Puncture or piercing

Many materials can suffer penetration from sharp objects. Again this can happen at any stage from basic material supply to the finished pack. Adequate cushioning and/or resistance to penetration helps to reduce the risks. Poor control of forklift trucks is a puncture hazard.

Climatic or environmental hazards

These may be ever-present hazards or hazards which are specific to a local environment. Although climatic conditions are covered by such phrases as arctic, antarctic, temperate, subtropical and tropical, severe conditions can occur elsewhere, i.e. in a deep freeze (–19 to –22°C), in a bathroom or kitchen where conditions can be worse than many tropical areas, displayed under high-wattage bulbs in a shop window, stored near pipes or heaters in a shop or warehouse etc. Climatic hazards therefore include the following.

Moisture

Moisture as liquid or water vapour may cause physical changes (e.g. dulling, softening, hardening etc.) or chemical change (hydrolysis, effervescence, etc.) It may also act as a carrier for other contaminants. Certain materials (including all plastics) are to some degree permeable to moisture, and even screw closures which appear to make a good seal are likely to permit some passage of moisture depending on the sealing medium, the torque, the evenness and shape of the sealing surface, the aperture size and circumferential area of the container etc. It must be emphasized that either moisture loss or moisture gain may be critical to some products.

Temperature

Extremes of temperature (cold and hot) or cycling temperature can cause deterioration to product and/or pack. Although higher temperatures generally represent an acceleration effect occasions can be found where deterioration increases at lower temperatures (certain plastic will become more brittle and crack, for example). A high temperature coupled with a high RH will produce a shower effect if the temperature is lowered sufficiently to reach dew point. Contamination from liquid moisture can then encourage mould and bacterial growth.

Pressure

Air-pressure differentials are frequently seen as a danger for materials sent by air using unpressurized aircraft. Pressurized aircraft are pressurized to the

equivalent to about 3000 m above sea level; hence there is a −0.25 bar differential compared to takeoff. Goods filled in factories at sea level and sent to mountainous areas, or vice versa, will suffer from similar patterns, i.e. goods packed in Johannesburg, South Africa, at 2000 m and then sent to Durban at sea level will be exposed both to a positive pressure and probably to a temperature change.

Light

Light consists of wavelengths from the UV zone through the visible to infrared. Although UV is a potential source of photochemical change, such changes may not always be visible. Printed or decorated packaging materials may also suffer from discolouration (white may go yellow, deeper colours may fade), and this may be seen as implying a change in product efficacy or strength. Although light can be excluded by using selected materials, tin plate, foil etc., opacity and/or colour may reduce penetration or filter out selected wavelengths. The additional use of UV absorbers in plastics may also restrict light rays entering the pack. It should also be noted that many products are protected by a carton, outer etc. for a larger proportion of their life. Protection may then only be necessary for a relatively short display or use period when exposure to light occurs.

Atmospheric gases

These include oxygen, carbon dioxide, nitrogen and any other airborne gases. Oxygen leading to oxidation is the more obvious hazard. Carbon dioxide, however, can cause a pH shift (unbuffered solution in plastic bottles, particularly LDPE, which is relatively permeable to CO_2) and/or lead to precipitation of some products. The permeation of the common gases through plastic is typically in the ratio of 1:4:20 for nitrogen, oxygen and carbon dioxide, respectively, the latter being the most permeable.

Odorous gases, or volatile ingredients associated with perfumes, flavours and product formulations may also pass into or out of a pack. If a volatile ingredient is lost from a flavour, an unpleasant odour or taste may result.

Solid airborne contamination (particulates)

Particulates may be carried by or in the atmosphere. In the case of most plastic contamination may be increased by electrostatic attraction under dry conditions whereby particulates are drawn from the atmosphere by electrical charges. The presence of particulates will inevitably increase microbiological contamination risk.

Biological hazards

Microbiological

There is a general tendency towards improved microbial control for all products. This means that the packaging materials must be reasonably clean initially and, when put together to form a finished pack, restrict any further contamination as much as possible. In the case of sterile products the pack and its closure must maintain a 100% effective seal against microbiological ingress, i.e. bacteria, moulds and yeasts. Ingress of yeasts is critical with sugar-based products, e.g. syrups, as fermentation may occur. Mould will also grow on cellulose-based materials, i.e. paper and board, if these are kept under humid conditions.

Other forms of infestation

In common with foods, other sources of infestation that can contaminate pharmaceutical products include attack by insects, termites, vermin, rodents or any other bird- or animal-contaminating source. Although this is more likely to happen under poorly controlled conditions of hygiene and housekeeping, such infestation can still occasionally cause problems even in the UK.

Pilferage and adulteration risks

Pilferage being a human failing is broadly another biological hazard. The example of the Tylenol poisonings, in 1982 has placed greater emphasis on the need for tamper-resistant packs. Prior to this, various seals were used to indicate whether any product had been removed or replaced, rather than as a means of protecting against deliberate adulteration. Security seals, a possibly preferred phrase to tamper evidence, are widely used for pharmaceutical products as a means of increasing and maintaining user confidence in the product and pack.

Chemical hazards

As chemical interaction, if inherent to the formulation, cannot normally be reduced or avoided by pack selection (unless it is associated with exchange between the product and external atmosphere), the main risk must relate to interaction or incompatibility between product and pack. Compatibility

investigations must basically cover any exchange that can occur between the product and the pack, and vice versa. These may be associated with interaction or contamination, covering migration, absorption, adsorption, extraction, corrosion, erosion etc, whereby ingredients may be lost or gained. Such exchange may be identifiable as organoleptic changes, increase in toxicity/irritancy, degradation, loss or gain of microbial effectiveness, precipitation, haze, turbidity, colour change, pH shift etc. Again other external influences may catalyse, induce or even nullify chemical changes.

Some examples of chemical interaction between a product and its pack and the resulting contamination are described below.

1. Adsorption of chemical entities on to component surfaces occasionally occurs. Losses of EDTA and certain preservatives (e.g. benzalkonium chloride, thiomersal and other mercurials) have been observed.
2. The more volatile preservatives, e.g. chlorbutol, phenol, 2-phenylethanol, show fairly rapid loss through low-density polythene by absorption and surface evaporation. If an external overwrap, which is not permeable to the preservative, is added, then loss can usually be restricted to relatively low levels, i.e. less than 10%.
3. Other surface-active ingredients which may be found in plastics may also enter the product by dissolution, surface abrasion etc. These include antistatic additives, slip additives, antislip additives, mould release agents, antiblock agents, lubricants etc.
4. Detachment of glass spicules may occur when alkaline solutions of citrates, tartrates, chlorides and salicylates are stored in soda glass containers. This may occasionally occur when treated or even neutral glass is autoclaved in the presence of similar alkaline salts.
5. Organoleptic changes may occur, caused by permeation of volatile or odorous substances through plastic materials, e.g. solvents from printing inks.

STAGES IN THE DEVELOPMENT OF A PACK–PRODUCT COMBINATION

Development stages

To follow the theme that no pharmaceutical excipient, drug entity, intermediate or finished product can exist without a pack means that knowledge of the pack employed is relevant for all developmental stages of a dosage form, from preformulation studies right through to the finally chosen pack. The placing of formulations on test without reference to the packaging material, the closing system, the torque to which a screw cap has to be closed etc. is condemned.

Attempts to accelerate deterioration by the use of excessively high temperatures should also not be employed. In the same way that products may change when exposed to accelerated conditions, packs and packaging materials may similarly suffer changes, e.g. caps can tighten and crack (if plastic) or loosen and become ineffective as closures. It is therefore advised that packs generally should not be exposed to temperatures of more than 45°C, and even at this condition the time period should not exceed 12 months, with a maximum of 6 months being preferable.

With modern analytical techniques, TGA, DTA, GC, GPC, DSC, HPLC, IR, UV, TLC etc., changes in product or pack are becoming easier to detect and define in terms of actual change. It is also interesting to note the change in emphasis from identifying product purity to the identification and quantification of impurity and degradation products which in earlier years would have remained unidentified. These can now be fully quantified by the analytical techniques available. However, the student is warned that total reliance on so-called scientific methods is not enough. Sensual observations related to feel, appearance, texture, colour, smell and taste (where safe) should also be used, as these simple observations can occasionally detect change before any analytical procedure has been developed.

The stages broadly associated with packaging development are as follows.

Preformulation

All preformulation studies need some form of container. It is therefore important to understand the limitations associated with any packaging contact material used to contain or retain the material under test even at this very early stage of product development.

Product formulation

Formulations and any intermediates all require to be contained and stored. It is therefore necessary to make certain that all packaging contact materials are defined and that all pack parameters (torque, heat seal etc.) are identified, controlled and documented

(all part of good laboratory practice (GLP) and good pharmaceutical manufacturing practice (GMP)) during formulation studies.

Consideration of container materials

It is important to have a basic knowledge of all packaging materials, their properties, characteristics etc., and the processes by which they are fabricated/decorated as a packaging container or component, as well as how these and any subsequent processes may affect their properties, e.g. sterilization by ethylene oxide can lead to ethylene oxide and ethylene glycol residues. Gamma irradiation of low-density polyethylene not only marginally reduces material flexibility owing to molecular cross-linkage, but can give rise to formic acid and formaldehyde residues.

Pack feasibility tests

This is the stage where a product (preferably the formulation selected for ultimate sale) is tested in a range of possible packs, usually over a range of conditions from say –20°C to 45°C, together with some cycling conditions covering a temperature–humidity range. Note the weakness of testing under constant temperature conditions, when in actual practice all conditions are cycling and therefore variable. In addition to the storage tests indicated above the immersion of pieces of pack, or pack components if plastic, in the product or a simulant – i.e. an extractive-type test – may also be employed. Extractive tests are usually mandatory for plastics used for injectables and ophthalmic products. All materials used for these tests should be given a provisional specification and thoroughly checked (more rigorous quality control tests) prior to any use. Feasibility tests usually extend over 1–12 months, with 3–6 months normally being the minimum before a decision to proceed with a certain pack is taken. However, decisions based on accelerated conditions are sometimes difficult to interpret, as limited failure at, say, 45°C or severe cycling conditions, e.g. 15–37°C, 50–90% relative humidity (RH) over 12-hourly cycles, does not necessarily mean a pack (or the product) will fail under more realistic climatic conditions.

Formal stability tests

Once sufficient confidence has been generated in the pack–product combination, formal stability tests, on which the shelf-life will be based, can proceed. Test conditions have been specified by the International Conference on Harmonization and adopted by the major regulatory bodies in Europe, the USA and Japan. Normally three large-scale batches of product in each pack variant are stored at 25°C and 60% RH for long-term stability purposes, and at 40°C and 75% RH for accelerated stability, and sampled over a period of 5 years at examination intervals of 0 (initial), 3, 6, 9, 12, 18, 24, 30, 36, 48 and 60 months. The data generated are sent to the regulatory authorities as part of the Marketing Authorization Application.

Ongoing stability

This consists of repeated stability on random batches from production in order to confirm that the shelf-life does not change.

Complaints

This is the final means of monitoring the success of the product and pack. It is somewhat similar to the monitoring and recording of adverse reactions in that it is a safeguard to both the company producing the drug and the person receiving it.

In all the above tests analytical and packaging technological support is essential to check both the product and the pack. Some of the types of test employed are identified in this chapter. The aims quoted earlier bear repeating as without them there is no assurance that a satisfactory product pack has been achieved, i.e.

1. a specifiable product;
2. a specifiable pack;
3. a specifiable series of processes controlling all operations, from raw material to the assembly of the finished pack, as this ultimately becomes the basis for the control of all future production.

To achieve these specifiable parameters requires the cooperation and coordination of virtually all company disciplines so that the final aim of having a successful product which is effective, safe and profitable is met. The pack must therefore carry the product image literally from inception to disposal.

PACK SELECTION

Factors influencing choice of pack

Before a pack can be selected and the relevance of the various hazards considered, it is necessary to establish a thorough background brief which covers

consideration of the product, the market, the distribution system, manufacturing facilities and other considerations. These are discussed below.

The product

Product detail must include chemical and physical characteristics of the drug entity, the excipients and the formulation. It must also cover any recognized routes of deterioration or degradation, the dosage and frequency of dosage, the mode of administration and type of patient (baby, child, teenager, adult, elderly, infirm etc.), any of which could influence the product and pack style and any controlling legislative detail. Whether the product is seasonal or has a year-round use may be a further influence on pack selection.

The market

The eventual channels of sale should be considered, i.e. where, when, how and by whom it is to be used or administered (e.g. doctor, dentist, nurse, patient etc.). Where the product may be used (i.e. clinic, home, hospital), whether for home trade and/or export, the quantity per pack and the predicted sales for initial launch and follow-up sales, must all be carefully considered during pack design and selection.

The distribution system

The distribution system should be carefully thought out, for example conventional wholesale/retail outlets, or direct to selected outlets. How a product is to be distributed may be more relevant to export markets where less sophisticated transport systems (mules, donkeys, camels etc.) may be used, and where climatic conditions may be more severe, particularly if intermediate storage facilities are non-existent (e.g. no dockside warehouses) so that the consignment will be directly exposed to the atmosphere.

Manufacturing facilities

The suitability of the manufacturing facilities may have to be considered for a number of reasons, i.e. new pack, increased sales, improvements in GMP, revised product, new product etc. For example, can the envisaged product be manufactured and packed with existing facilities, or are new plant, equipment and buildings required? If the latter applies then the likely profit must be balanced against the likely cost of the venture. For example, an entirely new facility for the introduction of a packed product by an aseptic or a terminal autoclaving process would be more expensive than that required for a solid or liquid dosage form.

PACKAGING MATERIALS

Packaging types, styles and systems can be broadly defined by the material of construction, e.g. glass, plastic, metal, rubber, paper etc., and the process used for fabrication, e.g. blow moulding, injection moulding etc.

Glass and glass containers

Glass has had a successful history for pharmaceutical products in that it offers transparency, sparkle, easy cleaning, effective closure and reclosure where applicable, high-speed handling, good rigidity and stackability, and depending on the selection of the correct type of glass, is generally inert. The two main disadvantages, fragility and heavy weight, have partially been reduced by surface coatings to increase the surface lubricity and careful design. The latter includes the avoidance of sharp angles and the use of adequate radii. Plastic-coated or plastic-sleeved containers are likely to allow even thinner glass containers to be produced. The specific gravity of glass normally lies between 2.25 and 2.5, whereas most plastics are well below 1.5 (PVC is 1.4–1.45).

Glass is basically of three types: neutral, (type I), surface-treated soda glass (type II) and soda or alkali glass (types III and IV). These materials may be converted into components by pressing, blowing into a mould, or the shaping of glass cane (tubular containers). Typical compositions of type I (neutral) and type III (soda or alkali) glass are as follows (Table 36.1).

Table 36.1 Typical glass compositions (%)

	Type I	Type III
Silica	66–74	66–75
Lime	1–5	6–12
Soda	7–10	12.5–19
Alumina	4–10	1–7
Boric oxide	9–11	—

In type I glass the alkaline element is largely eliminated by the use of boric oxide to neutralize the oxides of potassium and sodium. Neutral glass has a higher melt temperature (around 1750°C) and a narrower working temperature range, which, together with the higher cost of boric oxide and the greater likelihood of imperfections, usually means a cost of two to three times that of soda glass for containers made by the blow-moulding process.

Surface-treated glass (type II) is made by treating the hot surface of type III (soda) glass with sulphur dioxide, ammonium sulphate or, in some countries, ammonium chloride. This neutralizes some of the surface alkali radicals, producing a more neutral surface. The process, which may also be referred to as sulphuring or sulphating, invariably leaves a hazy surface bloom (normally sodium sulphate) and so washing is essential prior to use. Soda glass (type III) is the most widely used material where extraction of alkali metal ions is not critical to the product. Type IV glass has a similar composition to type III, but it cannot be guaranteed to have the same quality. All glass types are available in clear (white flint) and amber, with other colours, including green, being produced as special makings. Although the majority of all glass containers are made by a blow-moulding process, vials, ampoules and cartridge tubes, produced by 'shaping' glass softened by heat from a length cut from glass cane, find wide application for pharmaceutical and cosmetic products.

Metal and metal containers

Metal, mainly as tin plate or aluminium, was at one time widely used for rigid containers for tablets, capsules, pastilles, powders and even liquid products. A significant part of this usage has been lost to other materials over the last 10 years. Light flexible gauges of metal (aluminium, tin and tin-coated lead) were widely used for collapsible tubes. Thin gauges of aluminium are widely used as foil in combination with other support or heat-sealing polymer films. Although all the above uses for metals still exist, there has been a gradual reduction in their use other than for foil and for aerosols. In addition to containers, tin plate, aluminium and aluminium alloy have been widely employed for ancillary components, including closures. Although the use of tin plate and certain types of aluminium screw closures has reduced, the special aluminium alloy developed for rolled-on and rolled-on pilfer-proof closures has remained in use owing mainly to recent events related to security and tamper evidence.

The production of collapsible tubes made by a process known as impact extrusion has been fairly static recently, first because of the growth in plastic tubes, and more recently because of the introduction of a multi-ply lamination, once known as 'Glaminates'. Metal collapsible tubes in the UK are largely made in aluminium (plain or lacquered) with a few in pure tin. As aluminium work hardens during impact extrusion (i.e. it becomes less flexible and more springy) aluminium tubes have to undergo an annealing process to ensure that the metal becomes flexible and capable of being shaped and folded. If any interaction between the product and metal is likely an internal protective lacquer, usually based on vinyl or epoxide resins, can be added. As metal tubes have a shoulder bearing an orifice (which takes a closure) and an open end (through which the contents are filled and after which the metal is folded and crimped), they have in effect two closures and two areas of seal. Whereas the dispensing end can be sealed, with a blind nozzle and/or a screw cap to make a good seal, the folded metal is less reliable in terms of searching or mobile products. The fold can therefore be improved by the addition of a latex or heat-seal band. Although the number of folds in the filling end can also be varied (e.g. saddleback (triple) fold, double fold etc.) the longer tube length required for the former (say + 9 mm) will add to the cost.

Tubes with elongated nozzles and a controlled orifice size are used for eye ointments. Nearly all caps on metal tubes are wadless, being plastic and moulded in polyethylene and polypropylene.

Plastics and plastic containers

Types and uses

Past years have seen a significant expansion in the use of plastics, from a few thermoset caps and the odd container (for menthol cones and shaving sticks) to a point where plastics have become a major packaging material. Plastics now used are related mainly to the thermoplastic resins. The most economic four are polyethylene (low, medium and high densities), polyvinyl chloride (unplasticized and plasticized), polypropylene (homopolymer and copolymer) and polystyrene (general purpose and impact modified). Other selected materials find specialized usages for containers, devices, components, plies or coatings. These include nylon (PA), arcylonitrile butadiene styrene (ABS), styrene acrylonitrile (SAN), polycarbonate (PC), polysulphone, polyvinylidene chloride (PVdC), polymonochlorotrifluoroethylene

(PCTFE), polyester (PET) and polytetrafluoroethylene (PTFE). Plastic resins or polymers offer many attributes, in choice of material and grade, processes of fabrication and decoration, a wide selection in design, and physical and chemical properties, all on an economical basis.

Disadvantages

In theory plastics appear to have certain disadvantages, e.g. possible extraction, interaction, adsorption, absorption, lightness and hence poor physical stability; all are permeable to some degree to moisture, oxygen, carbon dioxide etc; and most exhibit electrostatic attraction, allow penetration of light rays unless pigmented black etc. It is necessary to be aware of other possible negative features. These include:

1. ***stress cracking***: a phenomenon related to low-density polythene and certain stress cracking agents such as wetting agents, detergents and some volatile oils;
2. ***panelling or cavitation***: whereby a container shows inward distortion or partial collapse owing to absorption of gases from the headspace, absorption causing swelling of the plastic, or dimpling following a steam autoclaving operation;
3. ***crazing***: a surface reticulation which can occur particularly with polystyrene and certain chemical substances (isopropyl myristate first causes crazing, which ultimately reaches a state of total embrittlement and disintegration);
4. ***poor key of print***: certain plastics, such as the polyolefins, need pretreating before ink will key. Additives that migrate to the surface of the plastic may also cause printing problems;
5. ***poor impact resistance***: both polystyrene and PVC have poor impact resistance. This can be improved by the inclusion of impact modifiers, such as rubber in the case of polystyrene and methyl methacrylate butadiene styrene for PVC. However, both increase the permeability of each.

The majority of these effects can be either overcome or minimized by one means or another. An industrial example will illustrate the point.

It was required to pack a nasal spray formulation in a plastic squeeze bottle which was available worldwide. This immediately called for a low-density polyethylene pack. The product, however, contained a volatile preservative system which both dissolved in LDPE and was lost from it by volatilization, thereby immediately suggesting that a conventional squeeze pack was unsuitable. The LDPE bottle was, however, enclosed in a PVC blister impermeable to the volatile preservative and fitted with a peelable foil lid (also impermeable). As a result of this combination the loss of preservative was restricted to less than 5% of the total, i.e. preservative soluble in the LDPE, and preservative in the air space of the PVC blister reached a point where equilibrium was achieved between product, LDPE and the surrounding air space.

Additives

Because plastics still tend to be seen as relatively new materials, those used for ophthalmic solutions and injectables have a specific extractives procedure to pass. However, a knowledge of the constituents that may be found in a plastic material is equally important. The constituents fall into four categories: the polymer, residues associated with the polymerization process, additives (those constituents added to modify the plastic in a specific way) and any processing aids (which are used to assist any part of the process). The list of residues, additives and processing aids varies according to the plastic involved. Natural polyethylenes usually are low in residues and are likely to contain only a small quantity of an antioxidant. Polyvinylchloride, on the other hand, invariably contains a stabilizer to restrict any degradation that may occur during the heat processing.

The residues, additives and processing aids that may be used, and therefore possibly extracted from, a plastic include the following:

Monomer residues	Modifiers
Catalysts	Emulsifiers
Accelerators	Antioxidants
Solvents	Mould-release agents
Extenders	Lubricants
Fillers	Stabilizers
Slip additives	Colourants – pigments and dyes
Antislip additives	Whiteners and opacifiers
Antistatic agents	UV absorbers
Antiblocking agents	Flame retardants
Plasticizers	Light excluders (e.g. carbon black)
Release agents	

Most plastics will include only a few of the constituents listed above. However, depending on the additive used, other properties of the plastic can be changed, e.g. fillers such as chalk or talc are likely to increase moisture permeation.

Fabrication of plastics

Unlike glass, plastic containers and components can be fabricated by a far greater number of processes. These include injection moulding, injection and extrusion blow moulding, injection stretch and extrusion stretch blow moulding, thermoforming, scrapless forming process (SFP), reaction injection moulding (RIM) and solid-phase pressure forming (SPPF), all of which relate mainly to thermoplastic resins. Designs, rate of moulding and cost all vary according to the process chosen and the number of moulds involved, i.e. single or multiple cavity. Irrespective of the process, all moulding operations operate to a 'cycle' whereby the basic resin is heated, softened, shaped in a mould or moulds, and cooled to a temperature at which the article can be handled without distortion. Virtually all plastics shrink in the moulding/cooling operation and allowance has to be made for this. After moulding plastics can be decorated or printed by another wide range of processes, i.e. silk screen, dry offset letterpress, hot die stamping, cliche or tampon printing, therimage, letraset, or labelled by heat-sensitive, self-adhesive labels or plain paper labels using a special adhesive. Knowledge of both the fabrication and decoration of plastics is essential when a plastic material is to be used in contact with a pharmaceutical product.

Paper and board

The use of paper-based materials (cellulose fibre) remains a significant part of pharmaceutical packaging in spite of the fact that paper is rarely used on its own for a primary pack. However, the list of paper usages covers labels, cartons, bags, outers, trays for shrink wraps, layer boards on pallets etc., and combinations of paper, plastic and foil which are discussed separately. Cartons are used for a high percentage of pharmaceutical products for a number of reasons, increasing display area, providing better stacking for display of stock items, and the collating of leaflets which would otherwise be difficult to attach to many containers. Cartons also provide physical protection, especially to items such as metal collapsible tubes. Cartons therefore tend to be a traditional part of pharmaceutical packaging. Fibreboard outers, either as solid or as corrugated board, also find substantial application for bulk shipments.

Regenerated cellulose film (trade names Cellophane and Rayophane) are still used as an overwrapping material either for individual cartons or to collate a number of cartons. However, it is being substantially replaced by orientated polypropylene film. Although paper, even when waxed, has relatively poor protective properties against moisture, both paper and board (ointment, pill and tablet boxes) were once widely used for primary packs, particularly for dispensing operations.

Films, foils and laminates

The development of plastic films (early 1950s onwards) as distinct from regenerated cellulose film based on viscose, and the process of laminating two or more plies selected from films, cellulose coatings, foil and paper, has seen an ever increasing use of the various combinations. These materials fall into different roles, such as supportive, barrier, heat seal and decorative. Paper, for example, is usually a supportive ply which can readily be printed to give decorative appeal. Aluminium foil, even in the thinnest gauges (particularly when laminated or coated with a plastic ply), offers the best barrier properties, which are not approached even by the most impermeable plastics. Metallization, a relatively new process whereby particles of metal are laid down on to a surface under vacuum, can significantly improve the barrier properties of a material but these do not approach the properties of pure foil. The reflective properties of both metallization and foil can add to the decorative appeal of a pack. Plastics, as either films or coatings, can be used for decoration, flexibility, to provide various barrier properties, heat sealability, see-through properties (i.e. transparency), and to protect the other plies within the lamination.

In terms of cost, paper/plastic (paper/LDPE or paper/PVdC) or single plies of coated regenerated cellulose film or coated polypropylene represent the more economical materials that can be used for strip packs, sachets, overwraps etc. Although in general costs increase as the number of plies increases, newer techniques such as coextrusion, where a number of plastic plies are extruded in combination, can produce complete laminations cheaper than those produced by individual bonding. However, lamination bonding is still essential for plies containing paper and foil, as these materials cannot be extruded.

Uses for films, foils laminations are numerous, e.g. sachets, diaphragm seals for bottles, strip packs, blister packs, liners for large containers, overwraps, flow wraps, and liners for boxes, either attached (e.g. Cekatainer and Hermetet cartons) or loose bag-in-box systems and bags. Each of the above is likely to use different materials or combinations of materials for a number of reasons. For example, a blister pack consists of a thermoformed tray with a lid made from

board, paper, foil or film with coatings, which will either tenaciously adhere to the tray and act as a push-through material or be peelable, so that the lid can be peeled back to gain access to the contents. The thermoformable portion, i.e. the tray, can again be made from a single material, e.g. polystyrene, polyvinyl chloride, polyester etc., or be a combination, e.g. PVC coated with PVdC, PVC/PVdC/PE/PVdC/PVC, PVC/PCTFE (i.e. Aclar).

The thermoforming operation, whereby a heated and softened plastic ply, or sheet is drawn into a cooled mould, can be done by vacuum, positive air pressure, mechanically by a die, or a combination of these. If thicker foil is incorporated into the basic web to give a combination of nylon or polypropylene/40–50 μm foil/PVC or polyethylene, then cold forming, whereby a web is stretched without perforating, can be carried out. Foil blisters, as these are known, when sealed with a foil lid, can provide a hermetic pack, i.e. one that excludes virtually any exchange of gases between the product and the surrounding atmosphere. Similar protection can be achieved by using a foil-bearing laminate for a strip pack. In this case either the foil laminate is stretched by the insertion of an item in the 'pocket area' when it is held against a recess in a heat-sealing roller, or the pocket area is prestretched prior to reaching the intermeshing heat-sealing roller position. In all cases the blister or pocket has to be especially designed for the item to be filled if the maximum economy of material and machine is to be achieved. The volume and area of a blister pack is very product/machine dependent, and although it may be practical to put an item in a larger size of blister or pocket (at additional cost in material and possible slower production speed), it is usually impossible to make packs smaller without technical risks being taken, e.g. product sticking to foil lid in a blister pack; product causing pocket to perforate in a strip pack etc.

Both blister and strip packs appear to offer a reasonable degree of child resistance, particularly if the materials are opaque (opinion based on actual recorded poisonings or accidents).

Rubber-based components

Rubber components may be made from either natural or synthetic sources. The majority of rubber usage is related to the closure of sterile products (aqueous or oil-based and freeze-dried or powdered solids). Although natural rubber is being replaced by synthetics it offers advantages in terms of resealing (multidose injections), fragmentation and coring (descriptions for the means by which particles are created when a needle is passed through a rubber), but is poorer in respect to ageing, multiple autoclaving, extractives, moisture and gas permeation and the absorption of preservative systems. Synthetic rubbers tend to reverse all of the above properties and some formulations actually contain natural rubber in order to improve resealability, fragmentation and coring. However, most rubber formulations are relatively complex and may contain one or more of the following: vulcanizing agents (many of which are sulphur based), accelerators, fillers, activators, pigments, antioxidants, lubricants, softeners or waxes.

The main types of rubber used for pharmaceutical products include natural rubber, neoprene, nitrile, butyl, chlorobutyl, bromobutyl and silicone. Of these silicone is the most expensive and, although the most inert, is readily permeable to moisture, gases and absorbent to certain preservatives.

As indicated above, rubber components are likely to contain more additives than plastics. They are therefore tested by basically similar extractives and product contact procedures before they are used for injectable or i.v.-type products.

Rubber gaskets are also found in aerosols and metered-dose pump systems.

CLOSURES

Functions of a closure

Closures may be required to perform any of the following functions:

1. To provide a totally hermetic seal. This is a closure that permits no exchange between the contents and the outside of the pack, e.g. a fused glass ampoule.
2. To provide an effective microbiological seal, e.g. rubber bung and metal overseal. As rubber is permeable to moisture and gases to some degree, a bacteriological seal may not be strictly hermetic.
3. To provide an effective seal which is acceptable to the product, i.e. a closure which is not hermetic or a total guarantee against bacterial ingress, but adequate for the product.

Providing a 'seal' frequently depends on the marriage of a hard material with a softer, more resilient one, so that the former makes a physical impression on the latter. Closures generally require consideration of the following:

1. To be resistant and compatible with the product and the product/air space. NB Product contact will vary according to how the pack is stood: upright, upside down, on side, intermittent contact during transportation, movement etc.;
2. If of a reclosable variety, to be readily openable and effectively resealed;
3. To be capable of high-speed application where necessary for automatic production without loss of seal efficiency;
4. To be decorative and of a shape that blends in with the main container;
5. To offer such additional functions as may be deemed necessary – to aid pouring, metering, administration, child resistance, tamper evidence etc.;
6. To prevent or limit exchange with the outside atmosphere, to a permissible level. As well as moisture exchange, this may have to cover gases, vapours and actual liquid seepage or leakage.

Although a closure can be affected by various basic means, i.e. adhesion, heat sealing, welding, crimping, mechanical impression, interlocking, stapling, sewing etc., the majority of systems are related to physical compression or heat sealing.

The ***physical compression systems*** include:

1. screw caps – in metal or plastic; prethreaded or rolled on, with or without a wadding system (i.e. wadless);
2. plug in – a friction push-in fit;
3. push over – where a flanged or raised ring portion is pressed over a bead or lip.

Some closure systems endeavour to combine one or more of these systems and thereby achieve a multiple seal: for example, a seal may press externally on a sealing surface and also on an internal bore. Wadless thermoplastic caps using a 'crab's claw' seal or a skirted bore seal are becoming increasingly popular.

Wadded screw caps either contain a wad plus a facing, a disc of resilient plastic, or have a flowed-in plastic compound. The wad may be of compocork, feltboard, pulpboard or expanded polyethylene, faced with such materials as aluminium foil, tin foil (expensive), polyethylene, a vinyl material or PVdC (Saran). The latter, which has good barrier properties and is reasonably inert, is now the most widely used. Foil or waxed foil is slightly preferable if a higher barrier material is required. The wadding materials mentioned above are occasionally used, usually waxed, on their own. Plasticized PVC, polyethylene or foamed polyethylene have also found selective usage. Flowed-in linings, although slightly inferior in barrier properties and inertness, offer production-line advantages in that there is no wad to fall out whereby a pack is left without an effective closure. Wadless caps also have this same advantage. The knowledge required to understand closure systems is frequently underrated.

Rolled-on (RO) and rolled-on pilfer-proof (ROPP) aluminium alloy metal caps have always been popular for the security of export products. The RO and ROPP closure consists of a plain metal shell containing a wadding or flowed-in system, which is placed over the container neck and top pressure applied to give a good impression on the wad. While the pressure is still held, the threads are formed by a mechanical inwards pressure. In the case of the pilfer-proof closure an additional perforated collar is rolled under a lower bead. This type of closure system is capable of maintaining an excellent seal and does not suffer from the occasional tearing of the wad facing that occurs when a conventional screw cap is applied to a substandard bottle finish. In the case of a rubber wad it also avoids any watch-spring affect. However, a RO or ROPP cap does require a slightly higher standard for the quality of the bottle neck.

Determination of closure efficiency

Closure efficiency, i.e. the ability to prevent undesirable exchanges between the product and the outside atmosphere, can be determined by numerous methods:

1. Placing a desiccant in a pack stored under high RH and detecting any moisture gain;
2. Putting liquid inside the pack, storing at high temperature and low RH, and then detecting any moisture loss as a reduction in weight;
3. Holding the empty pack under water, applying a vacuum and observing for leakage or liquid ingress. Adding a dye and a wetting agent to the water may assist the defection of leaks;
4. Putting liquid in the pack, inverting and applying a vacuum. A poor seal is detected by liquid seeping or leaking out;
5. Checking that cap-removal torque (assumes quality of bottle and cap) is satisfactory. Torque can be time–temperature related, particularly on plastic bottles, and so measurement should be made against a standard condition and test period;
6. Checking on compression 'ring' seal in cap liner when the system contains a liner or lining

compound. If indentation by bottle surface on liner is not uniform or continuous, a faulty seal can occur. This can be confirmed by 'painting' the bottle rim before applying the cap to the specified torque; if the liner does not have a continuous ring of the marker paint then closure is incomplete.

The above tests cover those carried out to establish that a closure is satisfactory in a development programme and those that can be used in production monitoring as part of online quality control. Some closuring systems using ratchet tamper evidence closures or rolled pilfer-proof seals have two release torques, one to release the cap from the bottle sealing surface and one to 'break' the tamper-evident feature.

Heat seals

As indicated earlier another widely used method of sealing is heat (direct and indirect). To achieve an effective and permanent heat seal the two sealants must be compatible (partly compatible sealants can be used to give a peelable seal). Four factors have to be controlled: temperature, pressure, dwell (the length of time that temperature and pressure are applied) and the cooling period. Contamination of any area of seal (for example with product) should be avoided, although certain plastics will seal better than others in the presence of a contaminant. The main heat sealants include polyethylene (low density), wax coatings, PVdC, Surlyn ionomer, selected vinyl-based products and certain types of modified polypropylene. Conditions of sealing vary according to the heat sealant, with the majority sealing between 75°C and 150°C. The total effectiveness of the seal is also dependent on other variables, such as seal width, seal pattern, pack shape, and the presence of creases or stress lines (particularly if product area is overfilled or undersize) etc. The seal is usually checked by vacuum (carried out under water, probably containing a dye plus a wetting agent) to see if ingress occurs, and by the force required to pull the seal apart.

Other sealing methods

Plastics can also be sealed by other techniques, such as ultrasonics, high-frequency welding, hot-air welding and by other heat sources (flame, infrared, induction etc.).

Suffice it to say that closure is a most essential part of both primary and secondary packs.

FILLING

Packaging lines

Products packed on production lines include such items as unit and multidose packs, closable and non-reclosable packs, sterile products produced aseptically or by terminal sterilization, or non-sterile products with or without a degree of microbiological control, preformed containers which have to be filled and sealed etc., or those packed by a form fill seal process etc. Packaging lines may be fairly conventional and involve unscrambling, cleaning, filling, closuring, labelling, cartoning (probably with leaflet insertion), outerization and finally palletization, or be selective to a specialist operation, e.g. blister and strip packaging. Additional operations that may be carried out on a production line include printing, batch coding, expiry dating, and the incorporation of administration aids etc. Glass bottles other than those used for the more critical types of product (injections, eye drops or similar sterile products) are normally manufactured and packed clean and then subjected to pressurized air and vacuum as a final cleaning process (for fibres) prior to filling.

The total efficiency and hence profitability of a packed product depends not only on the type of pack and the material selected but also on the production line operation and the packaging equipment chosen. For example, the filling speed for a tablet will depend on its characteristics: the size, shape, fragility, resistance to powdering if uncoated, surface lubricity etc., and the type of pack chosen. For a non-fragile easy-flowing tablet, filling speeds for a wide-mouthed container (100s) could be 20 000 tablets per minute, with 5000 for a blister pack and 2000 for a strip pack. Choosing a narrow-mouthed container irrespective of the material used could reduce the filling speed to below 10 000 tablets per minute.

Organization of packaging lines

The organization of a production line involves several factors which may ultimately affect output. These include labour, planned maintenance, staff training, online quality control (QC), facilities for batch coding and expiry dating where relevant, constant delivery of an adequate supply of materials to the agreed specification on to the line, and removal of the finished product from the line. Also required are clearly defined procedures for cleaning, start-up and close-down of the line, plus full documentation on both procedures and materials to be

handled (including reclamation, verification, reconciliation) and environmental control, which all form part of good manufacturing practice (GMP), and of course the actions to take should one piece or part of a machine break down. On a line performing a number of separate functions (i.e. unscrambling, feeding, filling, closing, labelling and cartoning) each following operation should be capable of higher output speeds than the former. Intermediate holding areas, such as a revolving table top, may also be placed between certain functions should, say, the capper unit temporarily need attention. In this way the first part of the filling operation can still continue provided the alteration or adjustment to the capper can be carried out relatively quickly. The rest of the line with the higher-output capability can then cope with the backlog of containers.

Packaging of certain products

The method of packaging obviously depends on the type of material to be packed. ***Solid items*** such as tablets and capsules are counted by, for example, resolving disc, slat counter, breaking a beam of light etc. ***Powder or granular products*** may be filled by volume using an auger (powder held in a rotating screw) or a filling cup, or by weight using a bulk feed plus trickle top-up. In the case of very small quantities a dosator which dips into a constant level of powder held in a reservoir can be employed. Whatever the fill process, observations have to be made as to whether it changes the characteristics of the powdered product, e.g. separation due to vibration, impaction due to compression etc. ***Cream and ointment-type products*** are filled volumetrically either by a piston-type filler or by an auger. ***Liquid products*** are also widely filled by a piston filler, a volume cup method using a pressure or gravity feed, or, where a rigid non-collapsible container is involved, by vacuum. In this process two tubes are lowered into the bottle and a seal made by a gasket on the container neck. When a vacuum is drawn on the vacuum tube, liquid flows through the filling tube until it reaches the level of the vacuum tube. The flow then cuts off. This process provides a clean fill, detects containers that are faulty (contain holes or have a dipped or uneven sealing surface), and of course operates a no-container no-fill feature. However, it is not ideally suited to frothing or more viscous liquids. A container that would collapse under a relatively low vacuum can be filled by vacuum, provided it is placed in an outer container and then sealed so that a vacuum can be drawn internally and externally to the pack.

Aerosols

Aerosols fall into a category of their own. They use a variety of materials (metal, glass, coated glass or plastic as containers, and combinations of metal, plastic and rubber for the valve) and offer a form of packaging where product and pack cannot be separated. In fact, an aerosol only comes into existence once it has been filled with a product and pressurized with a propellant. The latter may be either an inert gas or a halogenated hydrocarbon. Pharmaceutical aerosols by and large can be divided into metered-dose aerosols, where the total volume usually does not exceed 50 mL, and topical applications with a volume of 100 mL plus.

Metered-dose aerosols, many of which are used to treat asthmatic or bronchial conditions, are required to produce a product with a fine particle size, i.e. significant part of the cloud below 7 μm. This can be achieved by using powder or liquid aerosols where break-up is assisted by the propellant formulation and the valve. Attempts to produce a fine particle cloud using inert gases with a break-up system have so far fallen short of the standard required. The recent criticisms of the use of fluorocarbon propellants and their possible attack on the ozone layer has led to considerable innovation, with trends towards separating the product from the pressurized part of the system. As a result a number of bag-in-can or piston systems have been developed. Although such systems will dispense a solid or liquid product they will now provide sufficient break-up to give a true 'aerosol type' of dispensing.

Parenterals

Parenterals can normally be divided into small volume (SVP), e.g. injections, and large volume (LVP), e.g. i.v. injections and dialysis solutions. Although the size of containers varies between SVP and LVP the range of materials used, i.e. glass, plastic, rubber, is common to each. The background required for the development of a parenteral shares the same phases as other products. However, greater emphasis has to be placed on the packaging material and the sterilization process, hence the points made below.

Glass Glass can be sterilized by dry heat, steam or ethylene oxide, but is discoloured by γ irradiation.

Rubber Rubber can be sterilized by steam ethylene oxide subject to adequate aeration to remove residues, but has to be very carefully checked if irradiation is used as unacceptable physical and

chemical changes may occur. Rubber generally will not withstand dry heat.

Plastics Only a few plastics will withstand dry heat.

Sterilization Sterilization is possibly by steam, ethylene oxide, γ-irradiation and a process not previously mentioned – accelerated electrons – which is in essence a milder form of γ-irradiation. This rather sweeping statement has to be supported by adequate testing, as each case must be considered on its own merits. In the case of ethylene oxide treatment, degassing may be a critical stage as residues of ethylene oxide, ethylene glycol (hydrolysed ethylene oxide) and epichlorhydrin (if chloride ions are present) are all toxic in nature. γ-radiation can cause physical (due to molecular crosslinking) and chemical changes to a plastic.

Labelling

If the pack is not already printed, labelling normally follows closuring. Although the preferred labelling system in Europe and the UK is reel-fed self-adhesive labels, with limited use of plain labels and adhesive and heat-seal labels, this is not necessarily the case in other countries. The USA, for example, is more orientated towards heat-sealing (reel-fed and cut singles) than any self-adhesive system.

Both self-adhesive and heat-seal labels contain constituents that may prove to be migratory when adhered to certain plastics. Inactivation of benzalkonium chloride when self-adhesive labels have been applied to LDPE has been recorded.

QUALITY CONTROL OF PACKAGING

In the UK pharmaceutical products are broadly controlled by guidelines related to good manufacturing practice (GMP) and good laboratory practice (GLP). These cover all stages in the discovery, development, production, testing and sale of a pharmaceutical entity, and the means of providing records and documentation. Many of the aspects are related to quality assurance and quality control. However, any approval of a new drug discovery must involve thorough attention to detail, with the effective recording of information through all initial evaluation stages of drug preformulation, formulation, clinical and safety evaluation, packaging development and formal stability, leading to a production/marketing operation. In terms of packaging this should mean that any 'container' used for excipients or drug entities is identified, recorded and cleared for use at all stages. This should also include the storage of intermediates, and all tests irrespective of whether they are investigational or formal in nature, coupled to the procedures that control the pack, e.g. a product is to be stored in a glass bottle with a plastic cap with a liner/facing.

1. Glass bottles should be described by type (amber/white flint), type of glass (I, II, III etc.) and closure type (e.g. screw neck), and cleared as meeting a provisional specification in terms of dimensions, quality etc.
2. The closure should be defined terms of material (black phenol formaldehyde and wadding, i.e. pulpboard faced with 20 gm^{-2} Saran) and identified and cleared in terms of material, dimensions, quality etc.

If these materials are then used for a test at least the on and off torques should be recorded within, say, 1 hour of application. The climatic conditions (temperature and RH) may in certain circumstances need recording, e.g. in the packing of a moisture-sensitive product. This information and the associated documentation – probably a laboratory notebook – is all part of GLP. The recording, examining and passing of bottles prior to acceptance for any type of test is part of quality control. To put testing into perspective, no test should be performed without good knowledge of the pack, the product, and how the two items were assembled. For example, moisture gain by a sensitive product coupled with the finding of loose caps immediately poses the questions 'Were the caps adequately tightened?', 'Are the bottles and caps satisfactory?'. These questions cannot be answered unless a disciplined approach is used and adequate detail recorded. This attitude should apply to all developmental work, as well as the more important items such as clinical trial supplies, human volunteer studies etc. Tests between products and packs which are carried out to define a suitable pack/product are usually termed either feasibility or investigational studies. Once this stage has been completed, coupled with the development of analytical methods and analytical data, usually from accelerated storage tests, there should be a high confidence that the pack product combination chosen has an acceptable shelf-life or stability profile. The final proof that this opinion is correct lies with the formal stability programme, where the product is produced in the pack to be sold (or a near replica of it), by the final product method, and placed on test for a period of up to 5 years. The storage conditions in such a programme may range

from 4 to 50°C, with challenges associated with low and high RH, light, etc. with analytical periods of 0 (initial), 3, 6, 9, 12, 18, 24, 30, 36, 48 and 60 months. Regulatory bodies dictate the number of batches to be put on test: usually a minimum of three production-scale batches.

'Analysis' normally involves purity, identification of degradation products, loss or gain of moisture if relevant, microbial levels, effectiveness of preservative system, any exchange, interaction, adsorption, absorption between product and pack, and last but not least assessment of appearance, flavour, smell etc., as these aspects are sometimes more readily quantified by an observer than by a pure chemical analytical method. Specific analytical methods are essential for the main drug entity.

Because certain drugs degrade according to an Arrhenius plot a range of test temperatures can be chosen, i.e. 4°C, 15°C, 25°C, 35°C (or 37°C), 45°C. The temperatures required are 25°C and 40°C to equate with temperate and tropical parts of the world. However, it should be emphasized that most medicines finish up in the bathroom or kitchen (preferably out of the reach of children), and so even home conditions can be particularly severe while the product is being used or simply stored. Although disposal of drugs is recommended for dispensed medicines after the course of treatment has been completed both these and OTC products are inevitably stored for longer periods.

The procedures and controls for drug development are both intense and rigorous, and the same type of control is maintained throughout production, marketing etc., finishing with the monitoring of any complaints and/or adverse reactions. Products once launched are also sampled, say, one batch in 50 and put on a further stability test – known as ongoing or existing product stability – to ensure that the shelf-life profile is maintained. Any change following launch to processes, pack or product is not only similarly monitored but needs further input from packaging expertise.

Packaging development, whether carried out by a formulator or a special packaging section, must be based on a thorough knowledge of all packaging materials, packaging processes, basic test procedures as applied to paper, plastic, glass, metal, laminations etc., and devise programmes that provide the level of confidence that is essential between product and pack. In many cases this will involve packs that act as devices, or separate devices and user/patient type tests designed to establish functional efficiency and/or what will go wrong under conditions of misuse.

The ultimate of all pack clearance procedures is a pack specification that becomes the lead document for purchase and clearance of future deliveries of packaging materials.

Whereas quality assurance is the establishing of procedures that maintain quality, quality control is the actual testing activity. For instance, incoming packaging materials are first examined as a bulk delivery, then sampled on a statistical sampling basis, and finally examined in terms of variables and attributes for critical, major and minor faults to agreed acceptable quality levels (AQLs). As examination covers dimensional, aesthetic and functional aspects plus identification, (particularly relevant with plastics), it ensures that the material specified has been received. As production lines become faster, stoppages due to repairs, malfunction etc. have become more critical as ineffective and inefficient production can significantly affect costs. As an alternative to the statistical sampling of deliveries arriving in house, two other options are:

1. a random sample taken at the point of manufacture which is isolated and identified so that it can be checked by the user;
2. purchasing on warranty certification, which confirms that the quality specified is met as per agreed statistical testing scheme operated by the manufacturer.

With many, items where a high quality of cleanliness is essential sampling of the bulk delivery may lead to an additional risk of particulate or microbial contamination. In such circumstances the alternative schemes indicated above may be used to maintain the integrity of the incoming stock. Inspection and quality control play a further role with ongoing stock inspection, control on production lines, i.e. of the packaged product and the monitoring of finished (saleable) warehoused stock. As mentioned earlier, the final success of any product and its pack can only in the long run be equated to sales and complaints.

Concluding comment

This introduction to the packaging of pharmaceuticals should serve to establish the broad indepth knowledge required by the packaging technologist. The fact that all products only have a shelf-life when packed emphasizes the importance of what has frequently been seen as a minor role in the past. The success of any product depends on an effective marriage between product and pack.

BIBLIOGRAPHY

Briston, J.H. (1983) *Plastics in Packaging*. Properties and Application, Institute of Packaging Melton Mowbray, UK.

Briston, J.H. (1983) *Plastics in Packaging*. Conversion Processes, Institute of Packaging, Melton Mowbray, UK.

Dean, D.A. (1984) Pharmaceutical Packaging–present and future trends. Mang Chem. Aerosol News, March 10th, pp 282–284.

Dean, D.A. (1992) Packaging of pharmaceuticals: packages and closures; Institute of Packaging Melton Mowbray.

Montresor, J.M., Mostyn, H.P. and Paine, F.A. (1983) *Packaging Evaluation*. Institute of Packaging. Melton Mowbray, UK.

Moody, B.E. (1977) *Packaging in Glass*. Hutchinson, London.

Oswin, C.R. (1983) *Packaging Life*. Institute of Packaging, Melton Mowbray, UK.

Remington's Pharmaceutical Sciences (latest edition), Lea and Febiger, New York.

Ross, C.F. (1983) Packaging of pharmaceutical products, sterilisation and safety. Institute of Packaging, Melton Mowbray, UK.

Swinbank, C. *Intermediate Bulk Containers*. Institute of Packaging, Stanmore, Middx.

British and European Pharmacopoeias packaging specifications

BP 2000, Appendix XIX Containers

A Introduction (Ph. Eur. 3.2)
B Glass containers for pharmaceutical use (Ph. Eur. 3.2.1)
C Plastic containers and closures (Ph. Eur. 3.2.2)
D Containers for blood and blood products (Ph. Eur. 3.2.3–3.2.8)
E Rubber closures for containers for aqueous parenteral preparations (Ph. Eur. 3.2.9)

BP 2000, Appendix XX Materials used for manufacture of containers (Ph. Eur. 3.1)

A Materials based on PVC (Ph. Eur. 3.1.1–3.1.2)
B Polyolefines (Ph. Eur. 3.1.3)
C Polyethylene (Ph. Eur. 3.1.4–3.1.5)
D Polypropylene for containers and closures for parenteral and ophthalmic preparations (Ph. Eur. 3.1.6)
F Silicone (Ph. Eur. 3.1.8–3.1.12)
G Plastic additives (Ph. Eur. 3.1.13)
H Rubber for closures for containers for aqueous parenteral preparations and powders for freeze-dried products (Ph. Eur. 3.1.12)

37
Pharmaceutical plant design

Michael Aulton, Andrew Twitchell

CHAPTER CONTENTS

This chapter comprises three separate sections relating to the facilities and equipment in which medicinal products are manufactured. The first part discusses manufacturing facilities and describes the selection, design and utilization of a manufacturing facility, and introduces the concept of process validation. The second describes the properties and selection of the materials used in the construction of pharmaceutical facilities. The final part describes the types, causes and prevention of chemical corrosion of pharmaceutical plant.

MANUFACTURING FACILITIES FOR PHARMACEUTICAL PRODUCTS

After a considerable investment of both time and money in developing and testing many new chemical entities and products, some will reach the stage where there is the requirement to manufacture the product in large quantities. This section summarizes some of the factors that need to be taken into consideration when building a facility for producing pharmaceutical products on a large scale. This is only intended as an overview and the reader is referred to the work of Cole (1998) for a more detailed description of the design of pharmaceutical production facilities.

Manufacturing site selection

Pharmaceutical manufacturing facilities require considerable capital expenditure and are likely to be used for many years. It is therefore necessary to think very carefully about where the facility is to be located, as well as how it is constructed. One important consideration is whether the company should develop a new site or refurbish an existing one. Upgrading or redesigning an existing site may often be a more cost-effective option, but may not always be feasible. For example, there may be insufficient space to expand an existing site. It may be 'outdated' and not suitable for upgrading, or the company may be introducing a new product type where there is no existing plant, or the existing site may not be in a suitable location.

Local planning regulations which may delay the development of an existing site also need to be considered, as even small delays in bringing a pharmaceutical product to market can have considerable cost implications.

Considerations when selecting a new site

Most major pharmaceutical companies are multinational and have numerous disciplines contributing to the production of the final product (e.g. chemical synthesis, pharmacological and toxicological testing, product development and product manufacture), which may be located in many different countries. Increasingly, however, there is a tendency to concentrate specific disciplines in a small number of sites, and this may dictate where the production site is located. If this is not the case, then the advantages and disadvantages of locating in different countries need to be considered. Many countries, particularly developing countries, offer incentives to pharmaceutical companies to attract them to build manufacturing facilities in particular areas. These may, for example, take the form of tax incentives, such as no local taxes for a number of years, free land or building grants. Other factors that vary considerably between different countries and which may influence where the site is located include labour costs and availability, energy and construction costs, environmental legislation, likely opposition from the local residents and local employment legislation.

The chosen location must also have an acceptable local infrastructure, including suitable road links, services such as electricity and water, and the availability of a suitably qualified workforce.

Design and utilization of manufacturing facilities

Aims of manufacturing facility design

The major requirements when designing and using manufacturing facilities should be to ensure the following:

1. *Acceptable product quality.*
   - Each unit dose should contain the correct component(s) in the correct concentration and have the desired release properties. There should be negligible cross-contamination between products or chemical contamination from operators. This is particularly important when using highly sensitizing materials (e.g. penicillins) and very potent substances (e.g. hormones and cytotoxic drugs). In these later cases 'dedicated' facilities should ideally be used.
   - The facility must be designed to maximize protection against the entry of insects/animals.

- Microbiological contamination must be acceptably low both to protect the patient and to ensure a satisfactory product shelf-life.
- The factory should be designed to ensure that no unauthorized personnel gain entry to processing or storage areas.
- Each product must be correctly packaged and labelled to ensure the patient takes the correct medication and that it reaches the patient in a satisfactory condition. If the product is not of the required quality this will have considerable cost implications – for example if the batch has to be destroyed. If the batch has to be recalled after distribution, owing to errors such as incorrect labelling, this is very costly both financially and in damage to the company's reputation.

2. *An acceptable working environment.*
   - It is the duty of the pharmaceutical company to ensure that the manufacturing area is designed and operated so that the exposure of personnel to drugs, solvents etc. is at an acceptably low level. This may be achieved in some circumstances by controlling the air flow and quality in the area, but with more toxic materials special breathing apparatus with appropriate filters may be necessary.
   - Comfortable working conditions should be provided for process operators wherever possible, bearing in mind any particular product requirements, such as the need to prepare a product at a low temperature or humidity.

3. *Manufacturing efficiency.* Correctly designed and located facilities within the manufacturing area will improve the efficiency of the manufacturing process and thus reduce the cost of product manufacture. Careful consideration needs to be given to all the different stages involved in the process and how material will be transferred between stages, so that the production facility layout is optimized. For example, the use of gravity feeding of material between stages might require the facility to be sited on different levels, whereas the use of pneumatic or vacuum systems allow the facility to be sited on one level.

Achievement of manufacturing aims

Environmental conditions Control of the air temperature, relative humidity and particulate and microbial content is important so that the aims of the manufacturing process can be met. The specific conditions required depend on the product being manufactured. The following are illustrative examples.

- For a tabletting operation, typical environmental conditions might be a temperature of 21°C, a relative humidity of 35–40% and an air filtration target of 95% removal of particles less than 5 μm. These figures are such because powder flow is influenced by both air temperature and, particularly, relative humidity.
- Hard gelatin capsule filling processes, on the other hand, are typically carried out at temperature of 24°C and a relative humidity of approximately 30%. These figures are necessary because capsule shells become brittle and crack if the relative humidity and temperature are too low, and become soft and sticky and tend to jam the capsule filler if the relative humidity and temperature are too high.
- Air relative humidity is also vital when manufacturing deliquescent products, where levels as low as 20% may be required.
- With clean preparation areas the air particulate level is very important and the filtration target may be for more than 99% removal of particles of 1 μm or less.

Typical 'comfortable' working conditions for operators are 18–24°C and 30–60% relative humidity.

The temperature of the air may be controlled either by passing it through a steam-based or electrical heating system, or over refrigerant coils. The relative humidity may be increased by spraying water (of small droplet size) into the airstream, or decreased by using refrigeration systems or desiccant wheels.

In order to maintain the required conditions in the manufacturing area, the stale air must be frequently 'replaced' with 'conditioned' air. The frequency at which this needs to occur will depend on factors such as the nature of the production process, the number of operators present and whether the room is 'in use'. It may vary from approximately 20 changes per hour (typical solid dose production) to 500 (rooms for sterile production).

Architectural considerations Manufacturing areas should be of a suitable size, construction and location to facilitate proper operation, cleaning and maintenance. Any pipework, light fittings or ventilation points should be designed and sited to avoid the creation of recesses which might collect 'dust' and are difficult to clean. If possible, they should be accessible from outside the manufacturing area for maintenance.

Interior surfaces (walls, floors and ceilings) should be smooth, free from cracks and open joints, should not shed particulate matter, and should permit easy and effective cleaning or disinfection. Lighting should be sufficient for easy reading of in-process control systems, equipment dials etc.

Validation of manufacturing facilities

Validation and why it is required

Detailed coverage of validation is beyond the scope of this chapter owing to its complexity and the constantly changing requirements of the regulatory authorities. This section therefore summarizes the scope of validation with respect to manufacturing facilities and explains why it is important. In order to be aware of the latest position, the reader should consult information provided by regulatory authorities or other relevant authoritative sources.

Validation is a component of cGMP (current good manufacturing practice). Numerous definitions exist, including that of the Food and Drug Administration (FDA, USA), which has defined process validation as 'establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality characteristics'. In other words, validation is proving that a process works.

Essentially, the purpose of validation is to ensure consistent product quality. A new product or an old product manufactured using a modified process or facility cannot be sold until the 'process' has been adequately validated.

The scope of validation

When manufacturing a product all aspects of the 'process' need to be validated. This includes bulk drugs, excipients, suppliers, analytical methods, computer systems and personnel, as well as manufacturing equipment and processes. It must be ensured that the manufacturing equipment complies with the intended specification and that all services and recording equipment function correctly. The manufacturing process must be robust and produce a product with consistent properties. This is usually confirmed by manufacturing three full-scale production batches under specified conditions. In order to minimize cross-contamination between batches, the processes used to clean all equipment in which the product comes into contact must also be validated.

Types of validation

Validation may be divided into four types.

Prospective validation This is where the validation programme is designed ***before*** the equipment or facility is used, or before the product manufactured by the process being validated is distributed. This type of validation is employed when producing the data on initial full-scale production batches required to obtain a regulatory authorization for marketing and manufacture of that product.

Retrospective validation This is achieved by reviewing historical manufacturing and testing data of products or processes which are already in operation and working well.

Concurrent validation This is ongoing prospective validation or ongoing review and evaluation of historical data.

Revalidation Revalidation involves repeating all or part of the validation if, for example, a process or facility is modified.

Cost/benefits of validation

The cost of validation is considerable, with significant resources including personnel (from product development, quality assurance and engineering departments) and materials being required. Inadequate validation may, however, lead to the rejection of, or withdrawal of, legal authorization to manufacture and market the product. In other circumstances it may lead to expensive product recalls. Other benefits of well designed validation programmes are likely to include a reduction in process support required, fewer batch failures, greater output and speeding-up of marketing authorizations.

Cleaning as an example of a validation process

It is clearly desirable that a pharmaceutical product contains only those components that are supposed to be present. As most manufacturing equipment will not be used solely in the production of a single product, there is the possibility that, without adequate cleaning procedures, cross-contamination between products may occur. Although this should be detected during quality control tests, and therefore not endanger the public, it will still lead to costly batch wastage.

Cleaning protocols must therefore be validated to ensure that they are suitable. Because equipment can never be 100% clean, the aim of a cleaning procedure is to minimize the possibility of ***significant*** cross-contamination between batches of different products.

A typical specification would be that the contaminant level in the product taken by a patient is not greater than a 1000th of its lowest daily therapeutic dose.

Once a piece of equipment has been cleaned following a documented procedure, it is 'analysed' to detect the level of any product residue remaining. This may be achieved, for example, by:

- swabbing the equipment over a 100 cm^2 area at positions likely to be contaminated and analysing the swabs;
- collecting and analysing rinsings from final cleaning water (e.g. for cream manufacturing vessels);
- producing a placebo batch of the product in the cleaned container.
- Visual and tactile inspection is also useful for detecting contamination (except for very potent drugs).

It is a good idea for the product selected to validate the cleaning procedure to be the one that is likely to be the most difficult to clean of those produced in the equipment. Care must also be taken to ensure that any cleaning materials used are pharmaceutically acceptable, and that they are present in suitably low amounts. Once a cleaning process has been validated, operators must ***always*** follow the designated procedure to ensure a consistently successful process.

MATERIALS OF FABRICATION

Desirable properties and selection

In selecting materials for the construction of satisfactory plant the pharmaceutical engineer encounters problems involving chemical, physical and economic factors. The following brief outline indicates something of the scope and limitations of the choice of materials available.

Chemical factors

Two aspects of chemical action must be considered under this heading, namely, the possible contamination of the ***product*** by the material of the plant, and any effect on the ***material of the plant*** by the drugs and chemicals being processed. The importance of the first of these becomes evident when it is realised that impurities often have considerable toxicological effects, and even trace amounts of impurity may initiate decomposition mechanisms in the product. An example of the latter is the inactivating effect of heavy metals on penicillins. The appearance of a product may also be affected by changes in colour due to contamination from the materials of the plant.

It should be remembered that contamination from some materials may be innocuous, the products being non-toxic.

Our increasing knowledge of materials of plant construction is assisting greatly in providing plant that will be resistant to attack and deterioration in use from the effects of acids, alkalis, oxidizing agents etc. Alloys having special physical and chemical properties have been developed, and materials such as plastics have been introduced to meet the problems encountered.

Physical factors

An ideal construction material would satisfy all the following criteria. In practice no material is ideal and the selection procedure must be based on compromise.

Strength Sufficient mechanical strength is an obvious necessity and will be suited to the size of the plant and the stresses to which it will be subjected.

Weight In most cases weight should be reduced to a minimum, other factors being satisfactory, and especially in plant that may have to be moved about from place to place.

Wearing qualities These are particularly important where there is a possibility of friction between moving parts, an extreme case being the material used for the grinding surfaces of size-reduction mills.

Ease of fabrication It must be possible to process the material in order to manufacture the various units of the plant. Properties that enable materials to be cast, welded, bent or machined are of prime importance.

Thermal expansion The design of plant may be greatly complicated by the use of material that has a high coefficient of expansion. This increases the stresses and the risk of fracture with temperature changes, and the temperature range over which the plant will be operated is likely to be considerably restricted.

Thermal conductivity In heat-transfer plant, such as steam or water-heated vessels, good thermal conductivity is desirable. It must be remembered, however, that the bulk of the resistance to heat transfer may lie in the boundary layer films (see Chapter 38).

Cleaning Smooth polished surfaces simplify cleaning processes, and materials that can be 'finished' with such surfaces are ideal when scrupulous cleanliness is necessary.

Sterilization Where sterility is essential the material should be capable of withstanding the necessary treatment, usually steam under pressure. This factor is to some extent bound up with the previous one, as cleaning is a normal preliminary to the sterilization of apparatus and plant equipment.

Transparency This *may* be a useful property where it is necessary to observe the process. Borosilicate glass is used in the construction of pharmaceutical plant.

Economic factors

Cost and maintenance of plant must, of course, be economic. Here the main concern is not simply to obtain the least costly material: better wearing qualities and lower maintenance may well mean that a higher initial cost is more economical in the long run.

MATERIALS USED IN FABRICATION

A brief description of materials which are suitable for the fabrication or construction of pharmaceutical apparatus and manufacturing plant is given below.

Metals

Ferrous metals – steels

The element iron is abundant in nature in the form of its ores and, despite the tendency for the metal to corrode, it is still widely used. Steel consists of iron with added carbon, which may exist in the free state (graphite) or in chemical combination as iron carbide, Fe_3C. The phase diagram for the iron/carbon system is complex. Iron can exist in two allotropic forms. At temperatures below 940°C the stable form is α-iron (ferrite) but the γ form (austenite) can persist at normal temperatures if other metals such as chromium and nickel are alloyed with the iron. When steel is in the molten condition, austenite is able to dissolve carbon. Various eutectics and solid solutions may form on cooling, giving rise to steels of differing properties. These properties depend on the carbon content and any heat treatment that the steel may receive.

Mild steel This is the commonest form of steel for general purposes and has a carbon content between 0.15 and 0.3%. It is ductile and can be welded to give structures of high mechanical strength, but its resistance to corrosion is poor.

Cast iron The effect of increasing the carbon content beyond 1.5% is to lower the melting point of iron so that it can easily be melted and cast into sand moulds to form objects of intricate shape. Cast iron is resistant to corrosion but it reacts with materials such as phenols to give coloured compounds. Most pharmaceuticals are required to pass a limit test for iron, and if they are handled in cast-iron vessels they may fail to comply. Such vessels may be lined with resistant materials to take advantage of the strength of cast iron while shielding the product from the metal.

Cast iron is hard and brittle and can be welded and machined only with great difficulty. The hardness and corrosion resistance can be increased by adding silicon, though such high-silicon iron is extremely brittle.

Stainless steels These steels contain a proportion of nickel and chromium which confer a high degree of corrosion resistance. The stainless steels can be easily fabricated and polished to a high mirror finish. They are extensively used in the food, pharmaceutical, cosmetic and fermentation industries, where their high cost can be fully justified.

The most common type is the so-called austenitic stainless steel, where the nickel and chromium content stabilizes austenite at normal room temperatures. As can be seen from the phase diagram (Fig. 37.1) an alloy containing 18% chromium and 8% nickel (which is more costly than chromium) is the most economical form of austenitic stainless steel. The steel is often stabilized by the addition of 1% titanium or niobium, which prevents 'weld decay'. This can result from the removal of chromium as chromium carbide along the line of any welding that might be performed. Such depleted steel is then prone to corrosion.

Stainless steel owes its resistance to a tenacious oxide layer that forms on the surface. Materials such as nitric acid, which are oxidizing agents, can be handled in it but chlorides can penetrate the film and stainless steel equipment should not be used for hydrochloric acid.

One minor objection to stainless steel is its low thermal conductivity compared to other metals, but this is not usually significant as the main resistance to heat transfer may reside in the boundary layers (see Chapter 38).

Stainless steels can be used for most pharmaceutical plant equipment. Small apparatus commonly made from stainless steel includes funnels, buckets and measuring vessels. Sinks and benchtops are also made of stainless steel where a smooth surface with high corrosion-resisting qualities is necessary.

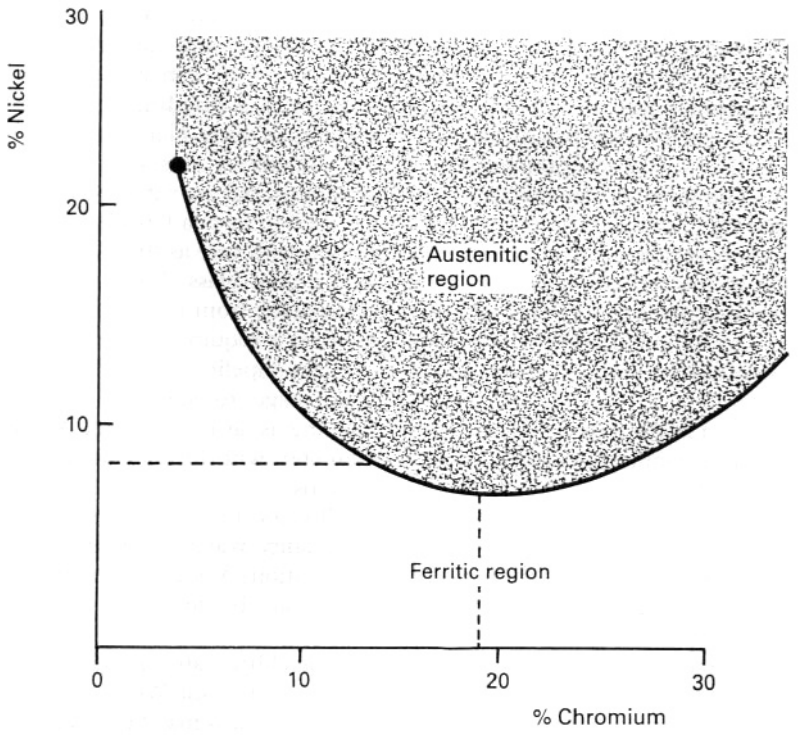


Fig. 37.1 Phase diagram for stainless steel.

The high cost often excludes its use, but often this may be justified even on a very large scale. For example, a great deal of the plant used in the production of penicillin is of this material, as it is by far the most satisfactory, being strong, corrosion-resistant, non-contaminating and readily cleaned and sterilized.

Non-ferrous metals

Copper Copper is malleable and ductile and thus easily fabricated. It has a thermal conductivity eight times greater than steel but is corroded by a number of substances, particularly oxidizing agents. It is attacked by nitric acid in all concentrations, by hot concentrated hydrochloric and sulphuric acids, and by some organic acids. Ammonia reacts with it readily to form blue cuproammonium compounds. Many drug constituents react with it, and for this reason copper is usually protected by a lining of tin when used for pharmaceutical plant. It was formerly used for evaporators and pans used for sugar coating (Chapter 28). Because of its susceptibility to attack by pharmaceutical materials there is a tendency today to replace it with stainless steel.

Copper piping is easy to make because of the ductility of the metal and it is used extensively for services such as cold water, gas, vacuum, and low-pressure steam. Where necessary, copper piping may be tinned, e.g. for distilled water, but nowadays stainless steel piping is used for this purpose.

Copper alloys These include alloys with zinc, tin, aluminium, silicon and nickel, all of which have their special uses.

Aluminium Aluminium has good corrosion resistance to many substances, although it is attacked by mineral acids (it is resistant to strong nitric acid), caustic alkalis, mercury and its salts. Its resistance is often due to the formation of a film on the surface. For example, acetic acid forms a film of gelatinous aluminium subacetate, which is then resistant to acetic acid, and pure strong ammonia solutions form a resistant film of aluminium hydroxide. It is also highly resistant to oxidizing conditions, as it forms a compact oxide film.

Pure aluminium is soft but more corrosion resistant than most of its alloys, such as Duralumin. Alloys combining corrosion resistance with strength are formed with the addition of small percentages of manganese, magnesium or silicon. Plant is easily fabricated and has excellent thermal conductivity.

For plant producing medicinal substances its most valuable property is probably the non-toxicity of its salts, which are, moreover, colourless. As it is also non-toxic to microorganisms it has considerable use in biosynthetic processes, such as the production of citric and gluconic acids, and of streptomycin by deep culture methods. The metal is also suitable for plant used for preparing culture media and for absorption and extraction vessels used in preparing antibiotics. As a result of the formation of resistant film it is used for acetic acid plant and storage vessels for ammonia. Because of its low density it is most useful for transport containers, such as drums and barrels.

Lead Much lead is used in the chemical industry because of its remarkable resistance to corrosion and the great ease with which it can be made into complicated shapes. The lead-chamber process for manufacturing sulphuric acid is one example from many in the heavy chemical industry. It is little used in pharmaceutical practice, however, because of the risk of contamination by traces of poisonous lead salts. Pharmaceutical materials must comply with a limit test for lead.

Tin Tin has a high resistance to a great variety of substances and, as its salts are non-toxic, it is widely used throughout the food industry. It is, however, weak, its main use being to provide a protective coating for steel, copper, brass etc. The coating of other metals with tin has been known for over 2000 years, and today more than half the output of the metal is used for this purpose. 'Tin plate' – sheet steel coated with tin – is used for containers.

Non-metals

Inorganics

Glass Glass is widely used in the laboratory and glass equipment can be obtained for manufacturing on a larger scale. Ordinary soda glass is used for bottles and other cheap articles, but is not satisfactory for large-scale plant or for containers where alkali contamination might be a serious drawback. For these purposes borosilicate glass is used, which has several advantages over soda glass. It is, for example, less brittle. It has a low thermal expansion and can be used safely over wide temperature ranges. However, its thermal conductivity is low and therefore it should be heated gradually to avoid fracturing. Pipeline may be used with pressures up to 8 bar if the pipe diameter is less than 50 mm, and pipelines with larger diameters are used with pressures up to 4 bar.

The special advantages of glass are that it can be easily cleaned and sterilized and, also, that the contents of vessels can be readily examined for colour and clarity. A disadvantage is the difficulty of joining sections of glass plant together. Ground-glass joints are sometimes satisfactory, especially in small-scale apparatus, but are rigid. Gaskets of rubber and plastic are used to form a flexible jointing, but these must be chosen with care as they are normally less resistant to attack than glass. For this reason gaskets are now often made from PTFE, but careful alignment of the sections is required for their successful use.

Glass pipeline is useful for transporting liquids from stage to stage in various operations. Such pipeline is available from 15 mm to 300 mm in diameter, with fittings for the assembly of complete systems.

All-glass stills are used for preparing pharmaceutical quality Water for Injections and other distilled preparations. Vessels up to 100 L are used, and larger tanks can be made by clamping glass plates in frames.

Glass fibres are excellent for heat insulation or refrigeration plant. Woven fibre may be used for filter cloths and glass may be sintered in the preparation of filters.

Fused silica Fused silica (Vitreosil) has an extremely low coefficient of thermal expansion and vessels made from it can be heated to red heat and plunged into cold water without breaking. It has a high melting point and it is difficult to fuse it completely, so that equipment tends to be opaque, although transparent forms are available.

Glass linings and coatings Metal may be coated with glass to give a protective lining. The dangers of such apparatus are those of uneven expansion of metal and glass, and of the glass surface accidentally chipping. Great care must therefore be taken in heating and cooling, and in protecting the glass lining from accidental damage.

Vessels of up to 50 000 L capacity, pipeline and fittings, valves, condensers, columns, pumps, stirrers and mixers are among the many glass-coated items still in use, although they have been largely replaced with stainless steel vessels.

Its high resistance to corrosion and ease of cleaning make glass valuable for pharmaceutical use.

Organics

Plastics The range of materials known collectively as plastics is so wide that only a brief account can be given. It is convenient to group them under the following headings:

- Rigid materials
- Flexible materials
- Coatings and linings
- Cements and filters
- Special cases.

'Keebush' is an example of a rigid material. This is a phenolic resin with various inert fillers selected for their particular purpose. It may be machined, welded, and worked in other ways, and is resistant to such an extent that it can be used for gears, bearings and similar items with a noise reduction of two-thirds compared to metal. It is resistant to corrosion, except that of oxidizing substances and strong alkalis. Any item may be made from this material – vessels, pipes, fittings, valves, pumps, fans, ducts, filter presses and many others.

Polyethylene and polyvinyl chloride (PVC) are similar materials and are either rigid or flexible, depending upon the amount of plasticizer added. They do not withstand high temperatures but are non-resistant only to strong oxidizing acids, halogens and organic solvents.

Rigid or semirigid mouldings may be used for tanks, pipes, ducts and other similar items; slightly flexible funnels, buckets and jugs are made which are almost unbreakable.

Polytetrafluoroethylene (PTFE, Teflon) is a semi-rigid plastic with extreme resistance to all agents other than fluorine or molten alkali metals. It is a slippery non-wettable material but can be bonded to metals as a protective coating. It is also usable at temperatures above 200°C, but it is costly, which prohibits its extensive use.

Metallic surfaces may be protected from corrosion by plastics of the polyethylene or PVC types prepared for coating with suitable plasticizers. Perfect adhesion of plastic to metal is sometimes difficult, and other disadvantages are the differences in thermal expansion of plastic and metal and the danger of the coating accidentally chipping.

Uses of these materials include the lining of tanks and vessels and the coatings on stirrers and fans. Plastic cements are used for spaces between acid-resistant tiles and bricks, and for similar purposes.

Special cases include transparent plastic guards for moving parts of machinery, and asepsis screens. Nylon and PVC fibres may be woven into filter cloths.

Rubber Hard rubber may be used for purposes similar to those mentioned for plastics. Soft rubber may be used for linings, coatings and valves. Rubber swells in contact with oils; it is subject to oxidation and is attacked by some organic solvents. Synthetic rubbers that have greater resistance have been developed.

CORROSION

General introduction and theory

Corrosion is a complex form of material deterioration that results from a reaction with its environment. It is the destruction of a material (usually a metal) by means other than mechanical. Most environments are corrosive. Corrosion is not restricted to reactions with strong acids. The following have been shown to cause corrosion:

- Air and moisture
- Rural, urban and industrial atmospheres
- Fresh, salt and distilled water
- Steam and gases, such as chlorine, ammonia, hydrogen sulphide, sulphur dioxide
- Mineral acids, such as hydrochloric acid, sulphuric acid, nitric acid
- Organic acids, such as acetic, citric, formic
- Alkalis
- Soils
- Organic solvents
- Vegetable and petroleum oils.

Importance of corrosion

Corrosion is very important in the chemical and pharmaceutical process industries, for a number of reasons. The possibility of any of the following occurring must be considered.

Safety

- Loss of pharmacologically active or toxic substances
- Loss of inflammable or explosive chemicals
- Sudden loss of material being processed at high temperature or high pressure
- Possibility of burns from acids etc.
- Corroding of equipment is known to have caused fairly harmless compounds to become toxic or explosive.

Financial losses

- Replacement of corroded parts
- Plant shut-down, i.e. stoppage due to unexpected corrosion failures. This leads to loss of revenue
- Loss of expensive chemicals.

Contamination of the product

This is of particular importance in the pharmaceutical industry as the product is usually for human consumption or administration. Metal ions can promote degradation reactions and can be toxic.

Appearance of plant

This is not a minor consideration. Satisfactory factory appearance is an important aspect of good manufacturing practice.

Corrosion mechanisms

The corrosive reaction of metals is generally electrochemical in nature. The relevant points of a simple, acceptable theory of electrochemical corrosion are summarized below.

1. An anode (–) and a cathode (+) form a cell in conjunction with an electrically conducting environment (electrolyte). Anodic and cathodic areas can arise on a single piece of metal because of local differences in the outside environment and in the metal itself. Different stresses, scratches and impurities produce varying electrical potentials. Cells with far greater electrical potentials are formed between different metals or alloys. Anodes and cathodes may be close (local cells) or far apart.
2. Direct current (DC) flows. The current is a flow of electrons within the metal(s) from anode to cathode. Metal ions (positive charge) travel from the anode in the electrolyte; they do not usually reach the cathode, but remain in solution.

Single-metal corrosion

Simple electrochemical corrosion is explained diagrammatically in Figure 37.2. This shows the mechanism for corrosion of a single piece of steel (predominantly iron) in water (dissociated by the presence of dissolved chemicals).

The anodic reaction is:

$$Fe \rightarrow Fe^{2+} + 2e^-$$

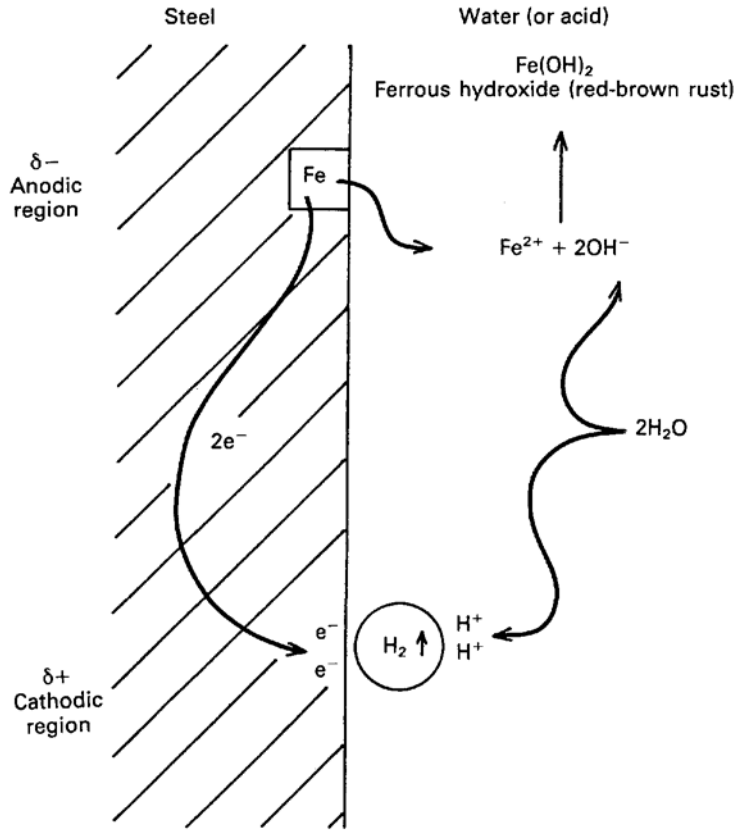


Fig. 37.2 Diagrammatic representation of the electrochemical corrosion of steel (iron) in water.

and the cathodic reaction is:

$$2H^+ + 2e^- \rightarrow H_2$$

The overall reaction is therefore:

$$Fe + 2H_2O \rightarrow Fe(OH)_2 + H_2$$

The important consequence of this reaction is that structurally strong metal leaves the anode as a metal ion in solution. The metal is therefore weakened and the solution contaminated with corrosion product.

Corrosion between metals

This is important because joints, flanges, seals, bolts etc. are often constructed of different metals. A cell is set up between the metals, one becoming the anode and the other the cathode. The potential difference between the metals depends on the electrode potentials of the metals. A knowledge of the electromotive series of metals is therefore important. These are summarized in Table 37.1, from which the following points can be noted:

1. A negative sign in the fourth column implies that the reaction in the direction of element to ion is spontaneous, i.e. it requires negative energy to proceed in that direction.
2. When two metals are coupled or in indirect contact with each other via the electrolyte a metal higher in the series will be the anode and one lower will act as the cathode.
3. Metals high in the table have less corrosion resistance. For example, metals above copper oxidize readily. This also explains why gold and silver are found as free metals in their native state, whereas aluminium and iron are found only combined as oxides.
4. The greater the difference in potential between two metals, the greater will be the rate of corrosion. For example, sodium reacts violently with water, iron slowly rusts, yet lead is not attacked by water.
5. When two metals are in contact, the corrosion rate of the anodic member of the couple is accelerated.
6. A metal will replace another metal in solution if the latter is lower in the series. This is best understood by examination of Figure 37.3.
7. The table therefore indicates whether or not a metal is attacked by acids or water, as metals above hydrogen are attacked because they replace hydrogen in solution,

   e.g. $Al + HCl(\text{in solution}) \rightarrow AlCl_3(\text{in solution})$

   $Fe + H_2O(\text{in solution}) \rightarrow Fe(OH)_2(\text{in solution})$

Table 37.1 The electromotive series of some commonly used metals

	Element	Ion produced	Standard electrode potential (V)
Active end	Na	Na^+	−2.71
	Mg	Mg^{2+}	−2.34
	Al	Al^{3+}	−1.67
	Zn	Zn^{2+}	−0.76
	Cr	Cr^{3+}	−0.71
	Fe	Fe^{3+}	−0.44
	Ni	Ni^{2+}	−0.25
	Sn	Sn^{2+}	−0.14
	Pb	Pb^{2+}	−0.13
Reference zero	H	H+	0.000
	Cu	Cu^{2+}	+0.35
	Cu	Cu^+	+0.52
	Ag	Ag^+	+0.80
	Pt	Pt^{2+}	+1.20
Noble end	Au	Au^{3+}	+1.42

Cathodic protection

This term is used to describe the technique whereby a structurally important metal is forced to become wholly cathodic (and therefore protected from corrosion) by attaching a more electronegative second metal to it. An example is galvanization, in which steel is coated with a layer of zinc. The zinc becomes the anode, thereby protecting the steel. The advantage of this technique over normal coatings is that even if the galvanized coating is scratched the exposed metal surface will remain cathodic. Metal objects can be protected in a similar way by attaching to them replaceable pieces of a more electronegative metal (e.g. a small aluminium block attached to a steel object).

Passivity

This is the phenomenon in which a metal appears in practice to be much less reactive than would be predicted by its position in the electromotive series. For example, one would expect aluminium to be extremely reactive, as its electrode potential is −1.67 V, yet it is commonly used in construction because of its lack of reactivity. The explanation is that aluminium does in fact react quickly with air, but an aluminium oxide coating is produced which is very resistant to further atmospheric attack.

Lead, which has a small electronegativity and would therefore be expected to react, at least with strong

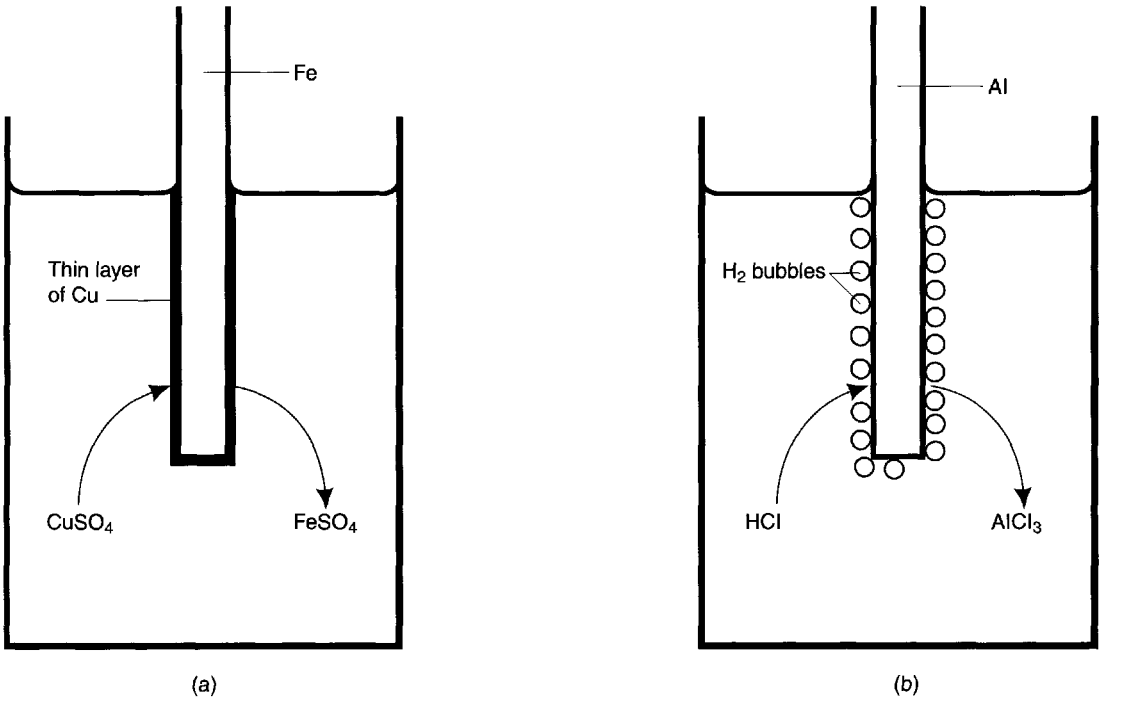


Fig. 37.3 (a) Reaction of iron in copper sulphate solution. The copper ions in solution are being replaced by iron ions (lower in table) at the same time as the copper from solution is being deposited as copper metal on the iron sheet. (b) Similarly, the hydrogen in the HCl solution is replaced by aluminium ions and the hydrogen is evolved as gas bubbles

inorganic acids, is in fact used in the production of sulphuric acid (lead-chamber process). In this example an impervious, unreactive layer of lead sulphate is formed on the outside of the lead sheets.

Types of corrosion

The various types of corrosion can be classified by the form in which they manifest themselves. The eight most common types of corrosion are:

1. Uniform attack or general, overall corrosion
2. Galvanic or two-metal corrosion
3. Concentration-cell corrosion
4. Pitting
5. Intergranular corrosion
6. Parting (or dezincification)
7. Stress corrosion
8. Erosion corrosion.

Uniform attack

This is the most common form of corrosion. It is normally an electrochemical reaction in which the anode and cathode move slowly over the surface of the metal. The metal becomes thinner and thinner and eventually fails, e.g. sheet-iron roof, zinc in acid. Uniform corrosion is the easiest to predict, discover and stop by means of the use of (a) more suitable materials, (b) inhibitors either in the metal or in the product, and (c) protective coatings (paint, plastic etc.).

Galvanic corrosion

This is corrosion between two dissimilar metals (see above). An example could be the case of a concentric pipe heat exchanger in which the main structure is of steel but the heat-exchange pipes are made of copper to improve heat transfer. In domestic central heating systems the commonest situation is to have steel radiators joined by copper piping.

Control of galvanic corrosion Use only one metal if this is possible. If not:

1. Select combinations of metals as close together as possible in the electrochemical series.
2. Avoid combinations where the area of the more active metal (i.e. the anode) is small, as this increases the anodic reaction rate. It is therefore better to have the more noble metal or alloy for fastenings, bolts etc.

3. Insulate dissimilar metals completely wherever possible.
4. Apply coatings such as paint, bitumen or plastic, but with caution. Small holes in the coating over an anodic region will result in rapid attack. Therefore, it is important to keep the coating in good repair.
5. Add chemical inhibitors to the corrosive solution. The nature of these inhibitors depends on the specific nature of the solution to be inhibited. Particular care must be taken in their selection when the corrosive solution is the pharmaceutical product, or is an intermediate for drug synthesis.
6. Avoid joining metals with threaded connections, as the threads will deteriorate excessively; spilled liquids or condensates can concentrate in thread grooves. Welded joints (using welds of the same alloy) are preferred.

Concentration-cell corrosion

Cells can form because of differences in the environment rather than differences in the metal itself. These are known as concentration (or solution) cells. There are two types, metal-ion cells and oxygen cells.

Metal-ion cells These can form in areas where stagnant liquid collects. Metal-ion concentration cells are caused, as their name suggests, by differences in metal-ion concentration in the corroding solution. A build-up of metal ions can occur in stagnant conditions caused by holes, gaskets, lap joints, surface deposits (scale, dirt) and crevices under bolt heads and rivet heads. They can also be caused by material, such as plastic, rubber etc. lying on the surface of the metal. Figure 37.4 shows the formation of a metal-ion concentration cell at a lap joint.

Oxygen cells These are similar to metal-ion cells in a number of ways, but here the corrosion is caused by differences in the dissolved oxygen content of the solution. The formation of an oxygen concentration cell at a lap joint is shown in Figure 37.5.

Control of concentration-cell corrosion This can be achieved in the following ways:

1. Use welded butt joints instead of riveted or bolted joints in new apparatus.
2. Existing lap joints should be welded, sealed or soldered to close the crevices.
3. Design vessels for complete drainage; avoid sharp corners and stagnant areas.
4. Inspect and clean deposits frequently.
5. Use solid, non-absorbent gaskets, such as Teflon, wherever possible.
6. Use welded pipes rather than the rolled-in type.

Pitting

This is a form of extremely localized attack where the anode remains as one spot. This results in rapid corrosion and deep penetration (small anode area, large cathode). Pits may be isolated or close together, the latter appearing as a rough surface. The pits usually occur at impurities in the metal, grain boundaries, small scratches, rough surfaces etc. Pitting is one of the most destructive forms of corrosion. It is difficult to detect in a laboratory corrosion test because the pits are very small and there will be little loss in weight.

Pitting is responsible for more ***unexpected*** plant equipment failures than any other form of corrosion. Additionally, the liquid in the pit becomes stagnant, resulting in metal-ion and/or oxygen concentration-cell corrosion. Furthermore, metal ions from the corrosion will accumulate in the pit. The process is therefore self-accelerating.

Control of pitting This is very difficult, but most of the points mentioned under concentration cells will help. If a test material shows the slightest signs of pitting in a laboratory test using microscopy, it must never be used.

Intergranular corrosion

Solid metals consist of a large number of grains (actually metal crystals) and thus have a granular (or polycrystalline) structure. Intergranular corrosion is localized attack at grain boundaries, with relatively little corrosion of the grains themselves. Because of stresses and structural imperfections at the grain boundaries (plus the increased concentration of impurities there), they are usually anodic. As corrosion proceeds the grains fall out and the metal or alloy disintegrates. This type of corrosion occurs in stainless steel, particularly after welding (known as ***weld decay***).

Control of intergranular corrosion This can be achieved in the following ways:

1. High-temperature post-weld heat treatment (solution quenching) can be used. This involves heating the metal to about 1000 K and then cooling it rapidly by quenching in water. This reduces the precipitation of carbon at the grain boundaries.
2. Add stabilizers to the metal, such as columbium, tantalum or titanium. These combine strongly

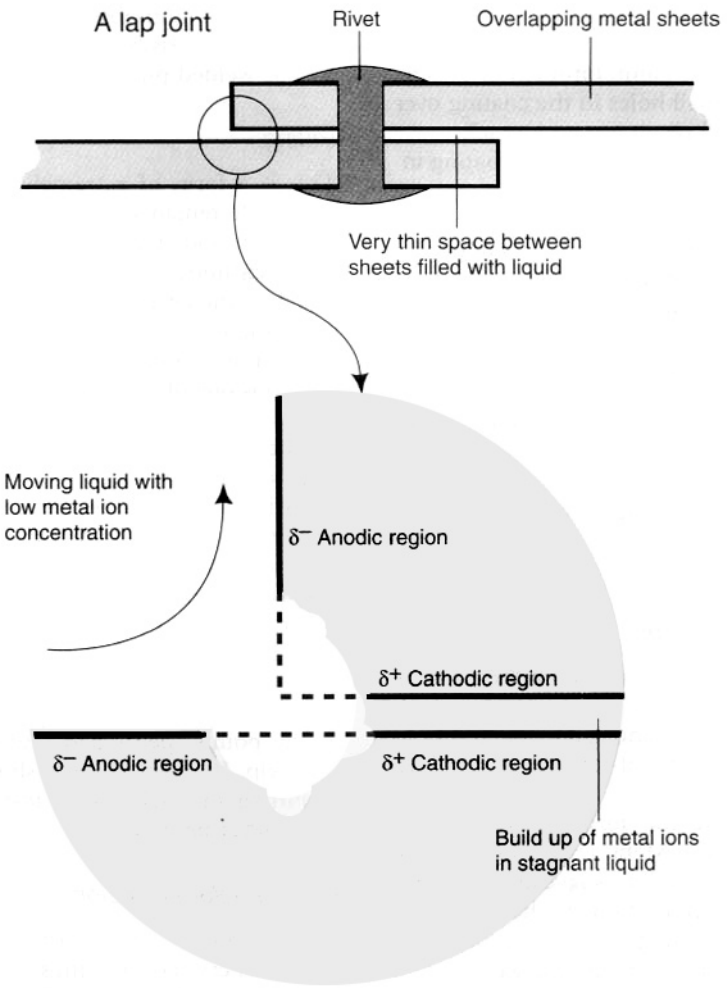


Fig. 37.4 A metal-ion concentration cell at a lap joint.

with free carbon to form carbides, leaving no free carbon at the grain boundaries.

3. Lower the carbon content of the steel to below 0.03%. Below this figure carbon is usually completely soluble.

Parting (or dezincification)

This is a general term referring to selective corrosion of one or more components from a solid solution alloy. The removal of zinc, particularly from brass (a 70/30 copper/zinc alloy), is common.

Control of parting This is not easy. Reduction of the corrosive environment and cathodic protection are suggested. Small amounts of arsenic, antimony or phosphorus can be used as 'inhibitors'. The addition of 1% tin to brass (to give Admiralty Brass) results in good resistance to seawater.

Stress corrosion

Generally a high stress in a piece of metal tends to make it more anodic. The greater the stress, the greater this effect. There are two main categories.

Stress-accelerated corrosion This is a decrease in corrosion resistance as a result of continuous static stress, such as applied stresses or residual stresses after welding (i.e. as occurs in pressure vessels).

Stress corrosion cracking This is an increase in the tendency of the metal to crack or show brittle fracture. It is usually caused by alternating stresses (i.e. vibrations).

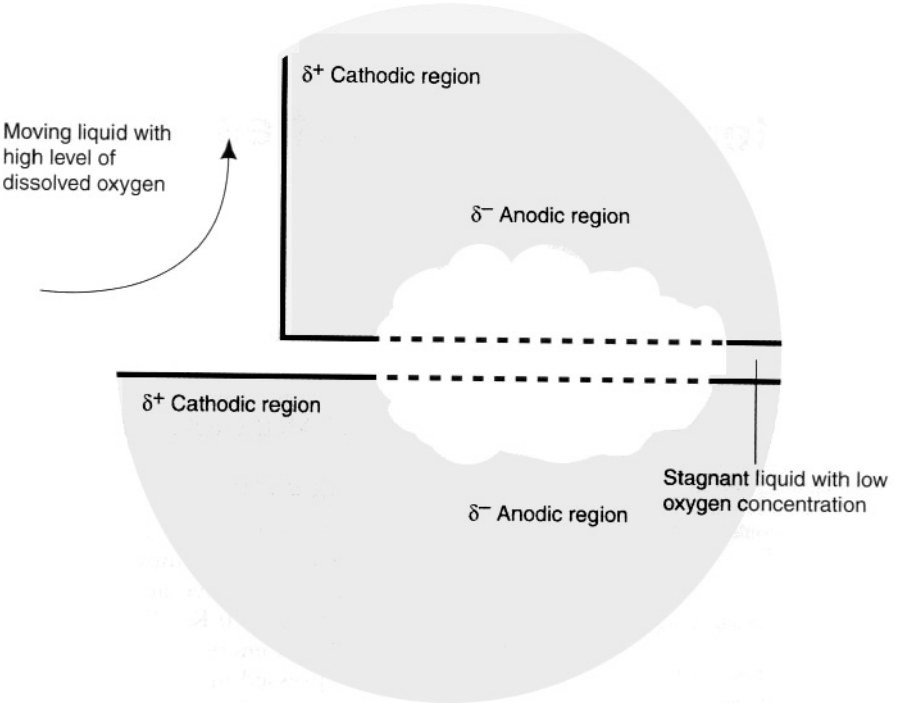


Fig. 37.5 An oxygen concentration cell at a lap joint.

Erosion corrosion

Erosion is a mechanical process, corrosion is electrochemical. They combine in situations where corrosive products which might have protected metals from further corrosion are eroded by mechanical wear or rapid fluid flow. This maintains a fresh metal surface in contact with the corrosive environment.

BIBLIOGRAPHY

Cole G.C. (1998) *Pharmaceutical Production Facilities – Design and Applications*. 2nd edn., Taylor and Francis. London.

Rules and Guidance for Pharmaceutical Manufacturers, current edition. HMSO 'The Orange Guide.'

38

Heat transfer and the properties and use of steam

Andrew Twitchell

CHAPTER CONTENTS

HEAT AND HEAT TRANSFER

Introduction

Heat is a form of energy associated with the disordered/chaotic movement of molecules/ions. A substance will have no heat content only if it is at absolute zero (0 K). Although heat is intangible, it is a form of energy and can be accurately measured and expressed in (preferably) joules (J) or alternatively Nm or kg m^2/s^2. The joule represents a very small quantity of heat; for example, 1 J will raise the temperature of 100 mL of water by only approximately 0.002 °C. A more practically useful quantity is the kilojoule (kJ) or the megajoule (MJ), which denote 1000 and 1 000 000 joules, respectively. It requires about 65 kJ of heat to raise 200 mL of water from room temperature to its boiling point. The temperature of a material is an indication of its internal energy: the greater the molecular motion the greater the internal energy and the higher the temperature. Heat transfer is the exchange or movement of heat energy and will occur spontaneously wherever there is a temperature gradient. The rate of heat transfer indicates how quickly heat is exchanged, and is expressed in J s^{-1} or watts (W).

Many pharmaceutical processes involve the transfer of heat energy. These include:

- melting materials and creating an elevated temperature during cream, suppository or ointment production;
- heating of solvents to hasten dissolution processes, e.g. dissolution of preservatives in the manufacture of solution products;
- sterilization of products, e.g. using steam in autoclaves;
- evaporation of liquids to concentrate products;
- heating or cooling of air in air-conditioning plant;
- drying granules for tablet production;

- heating air to facilitate coating processes;
- controlled cooling during cream manufacture.

It is therefore important to have a basic understanding of how materials may be heated and what influences the rate at which they can be heated (so that the heating or cooling can be controlled). This chapter considers the methods by which heat can be transferred and the factors that influence the rate of heat transfer. Particular emphasis will be given to the properties and use of steam, as heat energy provided by steam is the main source of heating in production processes. It is only intended as an introduction to this complex subject; the reader is referred to texts by Arpaci et al (1999), Long (1999) and Mills (1999) for more detailed information.

Methods of heat transfer

Conduction

Heat transfer by conduction in solids results from the movement of heat energy to adjacent molecules by vibrational energy transfer and the motion of free electrons. The molecule/electron donating the heat energy will subsequently vibrate to a lesser extent and therefore cool down, whereas the molecule receiving the heat energy will vibrate to a greater extent and therefore increase in temperature. No appreciable movement of molecules occurs. Heat transfer due to electron movement is generally a greater effect than that due to vibration of the atomic lattice, therefore metallic solids are good conductors and non-metallic solids are not (as they contain few free electrons).

In ***static*** fluids (and therefore through boundary layers; see Chapter 4) heat is transferred between molecules as a result of molecular collisions. When a fast-moving molecule from a region of high energy (or temperature) collides with slower-moving molecules from a region of lower energy, energy transfer takes place. Molecules with lower energy gain energy from the high-energy molecules, and the higher-energy molecules lose energy. As a consequence, the temperature of the initially high-energy molecules falls and that of the low-energy molecules rises.

Gases become better conductors at higher temperatures owing to the faster movement of the molecules, whereas most liquids (with the notable exception of water) become poorer conductors at higher temperatures.

Heat transfer by conduction is the main way in which heat is transferred through solids or fluid boundary layers, and is slow compared with heat transfer by convection.

Convection

Heat transfer by convection is due to the movement of molecules and their associated heat energy on a macroscopic scale. It involves the mixing of molecules and occurs within fluids, where the molecules are free to move.

Heat transfer by natural convection occurs when there is a difference in density within the fluid arising from the greater expansion and hence the lower density of the hotter region. Convection currents are set up as the warm, less dense fluid rises and mixes with colder fluid. As the molecules move, heat transfer ***between*** molecules can occur by conduction.

Heat transfer by forced convection occurs when the fluid is ***forced*** to move, for example by the movement of a mixer blade, or disruption caused by baffles. Heat transfer usually occurs more quickly by forced convection than by natural convection, owing to the greater intensity of movement and therefore the increased velocity of the fluid. If forced convection also induces turbulent flow (see Chapter 4) then the heat transfer process is aided, as there will be a reduction in the fluid boundary layer thickness.

Radiation

The energy emitted by the sun is transmitted in the form of electromagnetic waves through empty space. These waves can be reflected, transmitted or absorbed. When they are absorbed by a body on which they fall energy reappears as heat and the body increases in temperature. All hot bodies radiate energy in this way, but heat transfer by radiation is only of practical importance to pharmaceutical processing during drying (see Chapter 26).

Heat transfer by conduction and practical heat transfer

Of the three mechanisms of heat transfer described above, the one that will be considered in most detail is heat transfer by conduction. Heat transfer by convection, although important in pharmaceutical processing, is a complex process that is not completely understood and is difficult to define mathematically. More importantly, as far as calculating the rate of heat transfer is concerned, heat transfer by convection is rarely the rate-limiting step in the heat transfer process. If natural convection is inadequate then forced convection can be induced, e.g. by the use of a stirrer.

Many of the principles to be discussed in this chapter can be illustrated by the operation of a laboratory 'water bath', as illustrated in Figure 38.1. Heat energy from the gas burner is transferred through the container wall to the water in the bath, which therefore increases in temperature until its boiling point is reached. The heat gained is referred to as ***sensible heat***, as it produces an appreciable rise in temperature and the change can be detected by the senses.

When the boiling point has been reached further heating generates steam without any increase in temperature. This heat gain by the steam is termed ***latent heat of evaporation*** and is used to change the water from liquid into vapour at constant temperature. The steam produced rises and contacts the dish wall, which initially is at room temperature. The steam condenses on the cool outer surface of the dish and, in doing so, gives up the latent heat it contains, forming a layer of condensate on the dish which runs down over the surface and drops back into the bath. Fresh condensate is continually formed to take its place, so that a layer of condensate will always be present. The latent heat that is liberated passes by conduction through the wall of the dish and into the contents to be heated. Heat is then transferred through the fluid by natural convection and conduction. The term 'water bath' is therefore a misnomer and the equipment is described more accurately as a 'steam bath'.

The steam functions as a heat transfer agent whereby the heat from the gas burner is transferred by the liberation of the latent heat into the liquid in the dish. There are advantages in this indirect heating in that the temperature can never exceed 100°C (at atmospheric pressure), and therefore there is less chance of localized overheating. In addition, because the steam circulates over the whole dish surface, heating is much more uniform than it would be if the dish were heated directly over the gas flame.

The insert in Figure 38.1 shows a section of the dish wall in greater detail. If this is considered to be rotated into a vertical position and straightened slightly it would appear as in Figure 38.2. A temperature drop occurs from the temperature of the condensing steam to the lower temperature of the liquid in the dish. If this liquid is assumed to be of a lower boiling point than water, then eventually it will boil at this lower constant temperature and the temperature gradients would appear as in Figure 38.2. T_S denotes the steam temperature, T_L the temperature of the boiling liquid and T_O and T_I are the temperatures of the outer and inner surfaces of the dish.

In many pharmaceutical processes it is important to know or control the rate at which heat can be transferred, i.e. the quantity of heat transferred in unit time. This must be carefully distinguished from the total quantity of heat that needs to be supplied. Consider heating a beaker of water using a Bunsen burner flame. Under a low flame it might take 20 minutes to boil, whereas using a full flame it may only take 5 minutes. If heat losses to the environment are neglected (i.e. the heat transfer process is assumed to be 100% efficient), and the initial water temperature was the same each time, the total quantity of heat required to boil the water would be the same in each case. The rate of heat transfer, however,

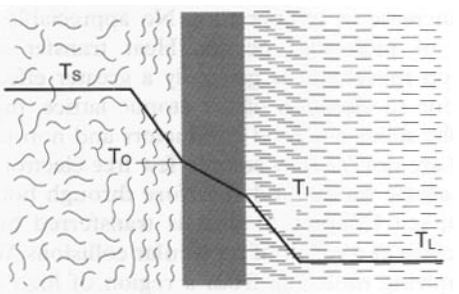


Fig. 38.2 Temperature gradients between steam and a boiling liquid.

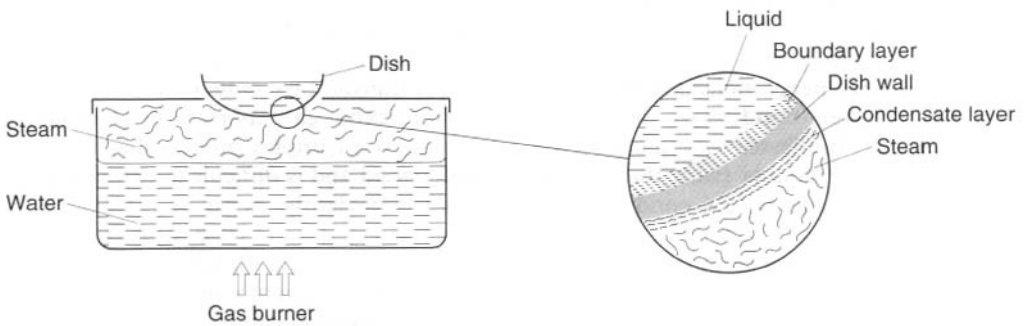


Fig. 38.1 Heating a liquid using a steam bath.

would be four times greater over the full flame – i.e. the water will boil in a quarter of the time.

In the heating situation depicted in Figure 38.1 there are three barriers to the flow of heat energy from the steam to the liquid bulk, namely the condensate layer, the dish wall and the liquid side boundary layer adjacent to the dish wall. The origin and nature of this type of boundary layer is discussed in Chapter 4.

The same quantity of heat must pass through each layer in turn at the same rate. Initially we will consider the factors that affect heat transfer through a single layer of material, in this case the dish wall. Conduction is governed by Fourier's law, which states that 'the rate of conduction is proportional to the area measured normal to the direction of heat flow and to the temperature gradient in the direction of heat flow'. The rate of heat transfer, i.e. the quantity of heat transferred (Q, joules) in unit time (t, seconds) will therefore depend on the difference in temperature (ΔT or $T_O - T_I$) between the outer and inner surfaces of the dish (the driving force for the heat transfer process), the dish thickness L_D and the area available for heat transfer, A. The proportionality constant is termed the thermal conductivity of the material and is denoted by the symbol K (K_D in the case of the dish). It gives an indication of the ability of the material to conduct heat: the higher the value the more easily heat is conducted.

Combining all these factors gives:

$$\frac{Q}{t} = \frac{K_D A(T_O - T_I)}{L_D} \qquad (38.1)$$

Equation 38.1 indicates that to increase the heat transfer rate (i.e. conduct heat more quickly) through a layer of material the value of A, ΔT or K may be increased or the value of L decreased.

By rearranging Eqn 38.1 and using appropriate SI units, thermal conductivity can be shown to have units of W/mK. Table 38.1 gives some typical representative thermal conductivity values and shows that metals are the best conductors, followed by non-metallic solids, liquids and gases. It should be noted that the K values will vary with temperature and also with the composition of the material (e.g. from 13 to 19 W/mK in the case of stainless steel).

Illustrative calculation for heat transfer by conduction
Q Calculate the quantity of heat passing in a period of 4 minutes through a stainless steel dish whose effective heating surface area is 25 cm^2 and whose thickness is 1 mm if the temperatures at the outer and inner surfaces of the dish are 90°C and 80°C, respectively.
A The first step is to convert all values into SI units. Thus:

Table 38.1 Thermal conductivity values of some materials encountered during pharmaceutical processing

Material	Thermal conductivity (W/m K)
Copper (pure)	386
Aluminium (pure)	204
Mild steel	43
Stainless steel (typical)	17
Glass	0.86
Water (at 20 °C)	0.60
Water (at 80 °C)	0.67
Boiler scale	0.09–2.3
Glass wool insulation	0.04
Air	0.03

$A = 25\ \text{cm}^2 = 25 \times 10^{-4}\ \text{m}^2$, $L_D = 1\ \text{mm} = 1 \times 10^{-3}\text{m}$, $K = 17$ W/mK (typical value for stainless steel; see Table 38.1), $T_O - T_I = 10$ K or 10°C (note because it is a temperature **difference** the numerical value is the same whether expressed in K or °C) and Q/t is expressed in watts.

Because $$\frac{Q}{t} = \frac{K_D A(T_O - T_I)}{L_D}$$

then $$\frac{Q}{t} = \frac{17 \times 25 \times 10^{-4} \times 10}{1 \times 10^{-3}}$$

$$= 425\ W$$

Therefore, 425 J of heat energy will be transferred every second, and in 4 minutes (240 s) the total heat transferred will be 425×240 J = 102 000 J = 102 kJ.

Heat transfer through multiple layers

In most practical circumstances heat needs to be transferred through multiple layers (as illustrated in Figure 38.1). In order to calculate the rate of heat transfer through more than one layer the thermal conductivity and thickness values of each layer need to be taken into account. It is not possible, however, to generate a value for the overall conductivity simply by adding the individual K values, as each layer offers a ***resistance*** to heat transfer. The overall thermal conductivity (K_O) of a system is inversely proportional to the overall thermal resistance (R_O), which in turn is the sum of the thermal resistance of the individual layers.

Thus

$$K_O = 1/R_O$$

and

$$R_O = R_1 + R_2 + R_3 + ... R_N.$$

This situation is analogous to the flow of electricity, where the total resistance to flow needs to be quantified in order to calculate the current flowing at any particular voltage.

To find the thermal resistance of an individual layer, Eqn 38.1 can be rearranged so that it becomes:

$$\frac{Q}{t} = \frac{\Delta T}{L / KA} \qquad (38.2)$$

This represents the general form of an equation for any rate process, where the rate at which the process occurs is expressed as a driving force divided by a resistance. In the case of heat transfer the driving force is the temperature difference across the layer and the group L/KA represents the thermal resistance of that layer. Other rate processes that can be expressed in this form include rate of filtration (see Chapter 22) and rate of dissolution (see Chapter 2).

Film heat-transfer coefficient

Referring back to the situation illustrated in Figure 38.1, it can be seen that heat has to pass three layers (condensate, C, the dish wall, D and the liquid boundary layer, L) in order to reach the liquid to be heated. The total thermal resistance for this heating process can be calculated by adding the thermal resistance of each layer (or film), i.e:

$$\text{Total thermal resistance} = L_C/K_CA + L_D/K_DA + L_L/K_LA$$
$$= 1/A\ (L_C/K_C + L_D/K_D + L_L/K_L)$$

as the area term is generally the same for each layer. Substituting this back into Eqn 38.2:

$$\frac{Q}{t} = \frac{A\Delta T}{L_C / K_C + L_D / K_D + L_L / K_L} \qquad (38.3)$$

where ΔT is the temperature difference across all the layers, i.e. the difference between the temperature of the steam and that of the boiling liquid.

Overall heat-transfer coefficient

Eqn 38.3 now accounts for the resistance of multiple layers and can be used to calculate the rate of heat transfer through layers of known thickness and thermal conductivity. It is useful for assessing the effect of individual layers on the overall heat transfer process, as indicated below.

The overall conductivity is represented by U, the overall heat transfer coefficient (OHTC), which gives an indication of the ability of a combination of layers to conduct heat. When heat has to be transferred through n layers, U can be calculated as:

$$U = 1/(L_1/K_1 + L_2/K_2 + L_3/K_3 + ... L_n/K_n) \qquad (38.4)$$

and thus for the situation shown in Figure 38.1:

$$U = 1/(L_C/K_C + L_D/K_D + L_L/K_L)$$

Substituting this into Eqn 38.3 gives:

$$Q/t = U A\ \Delta T \qquad (38.5)$$

U has units of W/m^2K and is only affected by factors that change the thermal conductivity or thickness of the layers through which heat is transferred; it is not affected by A. The value of the OHTC provides a useful indication of the overall conductivity of a heating system and is obtained using practically obtained data (see calculation number 6 at the end of this chapter), as it is not possible to determine the thermal conductivity and/or thickness of all the layers through which heat has to be transferred.

STEAM AS A HEATING MEDIUM

In pharmaceutical processes at anything other than laboratory scale, the most commonly used heating medium is steam. Steam is also very important as a sterilizing medium. The reasons for the widespread use of steam include:

- The raw material (water) is cheap and plentiful.
- It is easy to generate, distribute and control.
- It is generally cheaper than viable alternative forms of heating, e.g. electricity.
- It is clean, odourless and tasteless, and accidental contamination of the product is less likely to be serious.
- It has a high heat content (in the form of latent heat) and can heat materials very quickly.
- The heat is given up at a constant temperature, which is useful in controlling heating processes and in sterilization.

One disadvantage of the use of steam is that it is used at pressures that are typically two to three times higher than atmospheric, and thus steam presents potential safety problems and necessitates the use of high-strength piping.

To appreciate why steam is used in pharmaceutical processing and the principles of heat transfer using steam it is necessary to consider how the steam

is produced, its heat content, and how the heat content varies with pressure and temperature.

Heat content of steam

Consider heating 1 kg of ice-cold water in a closed container at atmospheric pressure. Initially all the heat supplied will be sensible heat, to raise the temperature of the water to the boiling point (100°C in this case). The quantity of heat (Q, joules) required to raise the temperature of a material can be calculated using Eqn 38.6:

$$Q = M\,S\,\Delta T \qquad (38.6)$$

where M is the mass heated (kg), S is the specific heat capacity of the material (J/kg K) and ΔT is the change in temperature (K or °C). The specific heat capacity is therefore the quantity of heat (J) required to raise 1 kg by 1°C or 1K.

For water at atmospheric pressure (1.013 bar/ 1.013×10^5 Pa), S = 4.2 kJ/kg K, and therefore the quantity of heat required to raise the temperature from 0°C to 100°C = $4200 \times 1 \times 100$ J = 420 000 J (420 kJ).

It should be noted that the value of S for water varies slightly with changes in temperature and pressure.

Once the water has reached boiling point further heat energy input will not raise the temperature of the water but will convert the boiling water at 100°C to vapour at 100°C, i.e. steam at 100°C. Steam at a temperature corresponding to the water boiling point at that pressure (as in this case) is referred to as ***saturated steam***.

The energy required to change unit mass from a liquid to a vapour at constant temperature is called the latent heat of vaporization (L, J/kg). For water at atmospheric pressure $L = 2.26 \times 10^6$ J/kg. L is not a constant value and depends on the steam pressure and temperature, e.g. L is 2.20 MJ/kg at 120°C and 2.14 MJ/kg at 140°C. The quantity of heat required to cause vaporization is calculated using the Eqn 38.7:

$$Q = M\,L \qquad (38.7)$$

where M is the mass vaporized (kg).

To convert 1 kg of water at 100°C to steam at 100°C, $Q = 1 \times 2.26 \times 10^6$ J (2.26 MJ), i.e. the steam now contains 2.26 MJ of latent heat energy. Steam in this state is referred to as ***dry saturated steam***, as *all* the water has been converted to steam. This form of steam should ideally be used for heating and sterilization processes, as it contains the maximum latent heat energy and no associated air or water.

If only half the water had been converted to steam the latent heat content would have been $0.5 \times 2.26 \times 10^6$ J = 1.13 MJ. Steam in this state is said to have a ***dryness fraction*** of 0.5, where dryness fraction is defined as the weight fraction of steam in a steam/water mixture. The dryness fraction is important because it governs the latent heat content of the steam, this being at a maximum when the dryness fraction = 1.

Once all the water has been converted to steam any further heat energy input increases the ***steam*** temperature, i.e. the steam gains sensible heat. Steam at a temperature above the saturation temperature is called ***superheated steam***. It only takes about 50 kJ of heat energy to raise the temperature of steam at 100°C at atmospheric pressure to steam at 200°C at atmospheric pressure.

The changes in the properties of water/steam with increasing heat input at atmospheric pressure are shown in Figure 38.3. The total heat content is measured from a datum at 0°C and is the sum of the sensible heat, latent heat and superheat.

If superheated steam at 200°C (A on Fig. 38.3) is cooled (e.g. if it contacts a colder surface) then it will lose heat energy. First, the steam will decrease in temperature until the small amount of superheat is given up and the steam reaches the saturation temperature. Only when the steam temperature has reduced to 100°C (B on Fig. 38.3) will the steam condense and the latent heat energy be released. When half the steam (0.5 kg) has condensed (C on Fig. 38.3), 1.13 MJ of latent heat energy will have been given up and the saturated steam will have a dryness fraction of 0.5. While there is still steam present, the temperature of the steam and any condensate formed that is in contact with the steam will remain at 100°C. Once all steam has condensed (D

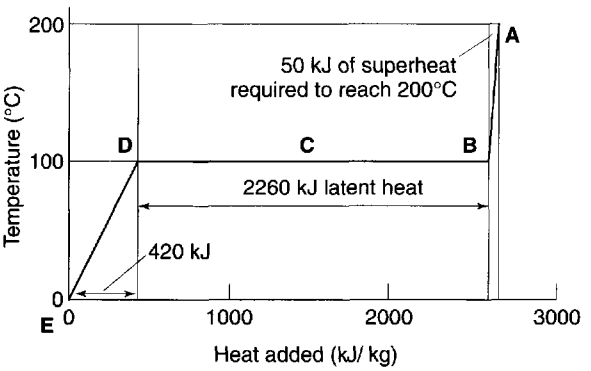


Fig. 38.3 Changes occurring on the addition of heat to ice-cold water at atmospheric pressure.

on Fig. 38.3) the condensate will lose sensible heat and decrease in temperature until the temperature gradient is reduced to zero. If the temperature of the surface in contact with the condensate is at 0°C then point E on Figure 38.3 will be reached.

It is important to note that most of the heat energy of steam (over 80%) is in the form of latent heat, and that when latent heat energy is released there is no drop in temperature. This latter point is important in sterilization processes and in maintaining temperature gradients when heating.

Effect of pressure on steam properties

The temperature at which water boils depends on the pressure exerted on the water surface. If the pressure is above atmospheric water will boil above 100°C, and if it is below atmospheric, for example if a vacuum is applied, water boils at a temperature below 100°C. The saturation temperature (and the temperature at which steam condenses) will also therefore be dependent on pressure. This is utilized in sterilization processes, where adjustment of the pressure allows selection of the temperature at which steam condenses and therefore the temperature at which the articles to be sterilized are exposed. Similarly, in heat transfer processes the desired temperature gradient can be achieved by adjusting the steam pressure. Some examples of how steam temperature changes with increasing pressure are given in Table 38.2.

Steam tables

The values in Table 38.2, along with temperatures at other steam pressures, can be found in ***steam tables***, as can values for the sensible and latent heat content at different pressures. It should be noted that there is not a linear relationship between steam pressure and steam temperature, and that the increase in steam temperature becomes less pronounced as the pressure increases.

Table 38.2 Relationship between steam pressure and steam temperature

Steam pressure (10^5 Pa)	Steam saturation temperature (°C)
1.013	100.0
2.000	120.4
3.000	133.7
4.000	143.8

Adverse effects of air in steam

There are two potential ways in which air may contaminate steam. First, water always contains dissolved air, and this air will be driven off when the water is converted to steam. Second, air will be present in equipment in the steam space when the process is started, and the incoming steam may not entirely flush this air out.

Air is a permanent gas which remains when the steam condenses to form a condensate layer, and thus forms an air layer in contact with the condensate.

The adverse effects caused by the contamination of air are twofold:

1. Air is a very poor conductor of heat (see Table 38.1) and forms a formidable barrier to heat flow. A very thin layer of air can markedly reduce the overall heat transfer coefficient (see calculation 4 at the end of this chapter), and the presence of as little as 1% of air in steam can result in a 50% reduction in the OHTC. Because the rate of heat transfer is proportional to the OHTC, there will be a corresponding reduction in the rate of heat transfer and thus an increase in heating-up times, process times and process costs.
2. Air will cause the steam temperature at any measured pressure to be lower than that of air-free steam. Steam containing air will not, therefore, be saturated. This arises from Dalton's law of partial pressures, which states that the total pressure in a system is the sum of the partial pressures of each of the components. Thus in an air/steam mixture the total pressure (P_T) is the sum of the partial pressure of the steam (P_S) and the partial pressure of the air (P_A). Therefore if P_T is 2×10^5 Pa and the steam contains 10% air, then P_A will be 0.2×10^5 Pa and P_S will be 1.8×10^5 Pa. The steam temperature, however, depends solely on the partial pressure exerted by the steam (not the total pressure), and will be 117.3°C (it would be 120.4°C if no air were present and the pressure were 2×10^5 Pa). Thus 10% of air has caused a reduction in the steam temperature at 2×10^5 Pa of 3.1°C. This is important in heating processes, where the temperature gradient will be lower than expected and thus the heating-up rate will be lower. NB: This effect is in addition to that caused by the poor thermal conductivity of air.

Both of the effects described above may also have potentially serious effects if air contaminates steam

in autoclaves. Its presence will give rise to an increase in the heating time required for the heating-up phase and the required sterilization temperature will not be reached. Sterilization processes use steam at a specific pressure with the implicit assumption that the steam will be saturated (i.e. no air is present) and will thus be at the saturation temperature. If this does not occur the material may not be exposed to a sufficient temperature for a sufficient time and the material may not be sterilized. Steam pressure alone should therefore never be used as an indirect measure of sterilization temperature.

Steam generation and use

Manufacturing installations for liquid and semisolid products

A diagrammatic representation of a jacketed installation typical of that used for the preparation of liquid and semisolid products is shown in Figure 38.4. This type of installation is available in sizes capable of manufacturing products from development scale (approximately 20 L) up to full production-scale batches of 20 000 L. They are constructed from a suitable grade of stainless steel, as this has acceptable thermal conductivity, is strong and easily fabricated, resists corrosion and is easily cleaned and sterilized (see Chapter 37). The outer surface of the jacket may be suitably lagged with materials having poor thermal conductivity, in order to protect the operators and reduce heat loss to the environment. In installations used for heating products using steam, the steam is usually generated in a remote boiler house that will supply steam to many different locations and pieces of equipment. The steam is usually generated at high pressures and temperatures (typically 6–8 × 10^5 Pa and 160–170°C), which enables it to be delivered to the equipment where required. Suitably strong pipes are used which need to be well lagged to avoid heat loss and condensation. The ancillary equipment shown in Figure 38.4 is discussed in more detail below.

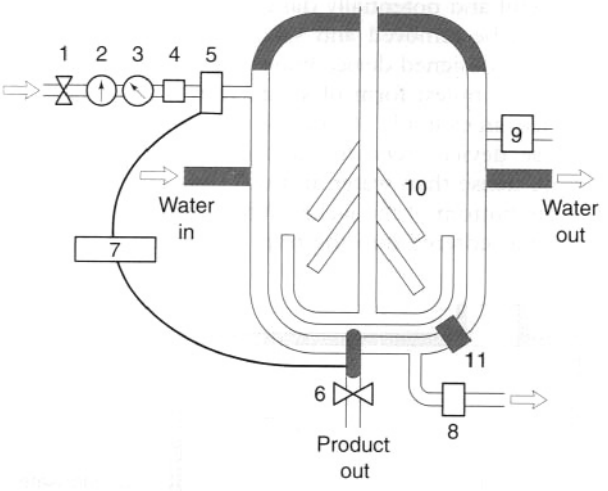


Fig. 38.4 Diagrammatic representation of an industrial steam-jacketed vessel. 1. Reducing valve; 2. pressure gauge; 3. temperature gauge; 4. safety valve; 5. control valve; 6. temperature probe; 7. temperature controller; 8. steam trap; 9. air vent; 10. stirrer; 11. homogenizer

Most heating installations use steam at about 1.5–3 × 10^5 Pa and a temperature of 110–135°C. This usually gives a sufficient temperature gradient to heat the product at the required rate, but reduces the chances of localized overheating. A ***reducing valve*** (1 on Fig. 38.4) is used to reduce and control/adjust the pressure to the desired level. This may be operated manually or, on larger equipment, controlled automatically via a ***control panel*** (7 on Fig. 38.4) and electronic signals from the ***pressure gauge*** (2 on Fig. 38.4).

A ***control valve*** (5 on Fig. 38.4) regulates the entry of steam into the jacket. As with the reducing valve, this may be either manually operated, for example on smaller development units, or automatically controlled so that the product is heated to the desired temperature and at the desired rate. An automatically operated system may function by selecting on the control panel the heating-up rate required, e.g. 2°C/min, and the final product temperature. A temperature probe in contact with the product sends a signal to the control panel, which in turn will cause the control valve to open if the product temperature is below the set value. When the control valve opens steam enters the jacket and the product starts to heat up. The panel will continuously monitor the product temperature, and the extent to which the control valve is opened will be controlled so that the product heats up at the required rate. When the desired temperature is reached a signal from the control panel closes the control valve and so no further steam enters. After a short time the product will start to cool slowly, as it is at a higher temperature than the surrounding environment. The temperature control system will detect this drop and reopen the control valve, so that more steam enters and the process is repeated. Using this type of control system the temperature of the product may be maintained within ± 2–3°C of the required value. Examples of where this may be used include maintaining a temperature of between 60 and 70°C during cream manufacture, or a temperature of 80–90°C when preparing

solution products where poorly soluble preservatives are employed.

The ***pressure*** and ***temperature gauges*** (2 and 3 on Fig. 38.4) allow the monitoring and recording of the steam properties and, as mentioned previously, the latter may be used to control the steam pressure used.

Because steam is generated at high pressures there is the potential to expose the equipment to pressures higher than it can safely withstand. To prevent this, a ***safety valve*** (4 on Fig. 38.4) is positioned before the control valve and is set to open and direct steam away from the installation if the pressure reaches a value in excess of that of the safe operating pressure of the system.

When steam first enters the jacket surrounding the product it contacts the cooler surface, condenses, and releases latent heat, which is then transferred through the various layers (condensate, vessel wall etc.) to the product. On condensing, steam contracts to a small volume (e.g. at 121°C, 850 mL of steam will condense to approximately 1 mL of condensate), which creates an area of lower pressure within the jacket into which more steam will then flow. Steam will therefore continue to enter the jacket to maintain the desired pressure until the product temperature reaches the steam temperature and no further condensation occurs.

If products are heated to temperatures above 60°C (as is often the case in cream or solution manufacture) then if left to cool naturally they will take a considerable time to cool to ambient temperature. This will be exacerbated as the volume in the vessel increases and if the vessel if efficiently lagged. Some products will need to be cool before volatile components (e.g. flavourings) are added, and to use the manufacturing vessels efficiently the cooling of these products will need to be accelerated. This can be achieved by circulating a cold fluid, e.g. water or 'brine' (a concentrated salt solution), through the vessel jacket. The latter can be at a temperature below 0°C and so will give a greater temperature gradient and faster cooling. If the rate of cooling is important (e.g. to avoid the formation of 'lumps' of higher melting-point components in cream and ointment manufacture) then the inlet and outlet of the cooling fluid can be controlled by a system similar to that used to control the heating rate.

The presence of a ***stirrer*** (10 in Fig. 38.4) helps ensure the product is evenly heated. The flow created will reduce the thickness of the boundary layer adjacent to the heating surface and thus speed up the heating process, and will mix the components of the product. If more intense mixing is required, as will be the case in the manufacture of emulsion, lotion and cream products, then a homogenizer (see Chapter 13) can be used. This will normally be sited at the bottom of the vessel, as shown in Figure 38.4. Where aqueous and oily phases need to be mixed when they are both at elevated temperatures, two jacketed vessels need to be sited close together and the appropriate phase pumped into the vessel containing the homogenizer.

The presence of a vessel lid protects the product from the operator and environment and vice versa. In addition, if the lid can be sealed then negative or positive pressures can be applied above the product's surface. A negative pressure (vacuum) is useful to minimize the incorporation of air during the manufacture of viscous products, especially when the homogenizer is used. This can avoid the manufacture of a product with an unsightly appearance and can reduce stability problems. A positive pressure can be used to aid emptying of the vessel.

Steam traps To ensure maximum heating efficiency the apparatus should minimize the amount of air and condensate present in the jacket. If the condensate is allowed to build up it will gradually reduce the area over which steam can condense, and therefore progressively slow down the heating process. If the condensate completely fills the jacket then heating will stop. The consequences of not removing the air from the jacket are described above. Fitting a simple drainage pipe to the jacket would be ineffective, as this, as well as removing condensate and air, would also allow steam to escape, which would be both wasteful and potentially dangerous. Condensate and air can be removed and steam retained by using a suitably designed device known as a ***steam trap***.

The simplest form of steam trap is a mechanical device, an example of which is shown in Figure 38.5. These devices rely on the fact that condensate is more dense than water and will thus tend to collect at the bottom of the jacket. When sufficient condensate has entered into the trap, the float will rise and

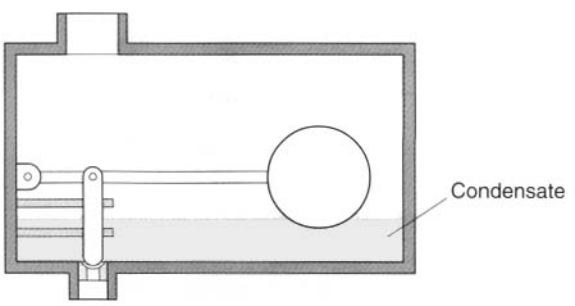


Fig. 38.5 Float-type mechanical steam trap.

open the outlet, allowing the condensate to drain away.

Mechanical traps are robust but have the major disadvantage that they do not allow air to escape, as there will always be condensate in contact with trap outlet. The alternative type of steam traps are referred to as thermostatic devices, and these rely on the fact that condensate can lose sensible heat and thus be at a temperature which is lower than the steam. A common form of thermostatic steam trap is the ***balanced pressure thermostatic steam trap*** shown diagrammatically in Figure 38.6. This contains a capsule in the form of a bellows, in which is a liquid having a boiling point a few degrees below that of water. Thus, when the capsule is surrounded by steam the liquid in the capsule boils and causes the bellows to expand and close the outlet. When condensate enters the trap it will lose sensible heat and decrease in temperature; the trap is usually constructed of a material with good thermal conductivity, e.g. copper, to hasten this heat loss. The condensate then cools the capsule, the liquid in the capsule ceases to boil (condenses) and the bellows contract, thereby opening the outlet and discharging the condensate. When the condensate is removed steam surrounds the bellows, causing the trap to close again. These traps will operate over a wide pressure range, as any increase in steam pressure not only raises the boiling point of water but, because the same pressure acts on the surface of the bellows, will elevate the boiling point of the liquid in the bellows by a similar amount. Hence the alternative title applied to this type – balanced pressure expansion trap – as it will always operate a few degrees below the saturation temperature of the steam.

Although these devices tend to be less robust than mechanical traps they have one major advantage in that they will also vent air from the jacket. Because air is more dense than steam, air will tend to collect at the bottom of the jacket and enter the steam trap. Contamination of steam with air will cause the steam temperature to be lower than the water boiling point at any operating pressure (see above), and once sufficient air is in the trap to reduce the steam temperature to below the boiling point of the liquid in the bellows, the bellows will contract and the air-contaminated steam will be removed.

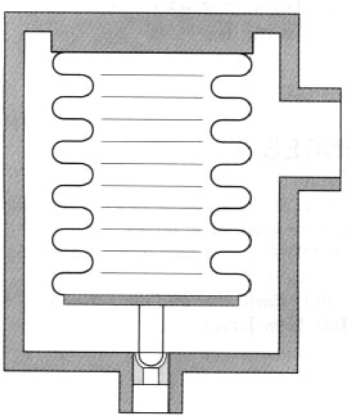


Fig. 38.6 Balanced pressure thermostatic steam trap.

Generally there are at least two traps on a heating installation, one at the bottom of the jacket to remove condensate and air generated during the heating process, and one towards the top of the jacket on the opposite side to which steam enters. This latter trap (which is of the thermostatic type) acts as an air vent and will be open when the equipment is started and therefore help to flush out the air initially present in the jacket. It will close when the temperature of the steam exiting the vent is sufficient to cause the liquid in the bellows to boil.

The condensate released from the jacket will possess significant heat energy and may be fed back to the steam boiler or used for other manufacturing plant, such as air-conditioning systems.

Example calculations involving the use of steam

This chapter concludes with some calculations that illustrate various points raised in the text.

1. **Q** What is the energy requirement to produce 13 kg of dry saturated steam at a pressure of 2×10^5 Pa (2 bar absolute/1 bar gauge) from water at 18.4°C? Assume the specific heat capacity of water is 4.21 kJ/kg K and the latent heat of evaporation is 2.20 J/kg.

   A The saturation temperature at 2×10^5 Pa is 120.4°C. The heat energy required to raise 13 kg of water to 120.4°C = $13 \times 4210 \times 102$ J = 5.582×10^6 J.
   The heat energy required to convert 13 kg of water at 120.4°C to steam at 120.4°C = $13 \times 2.2 \times 10^6$ J = 2.86×10^7 J.
   The total heat energy required therefore = $5.582 \times 10^6 + 2.86 \times 10^7 = 3.418 \times 10^7$ J (34.18 MJ).

2. **Q** What temperature would be reached if the steam produced in Question 1 was used in a steam-jacketed vessel to heat 150 kg of water whose initial temperature was 21.8°C? Assume there are no heat losses to the environment.

A The 2.86×10^7 J of heat energy required to convert water to steam will be released as latent heat when the steam condenses during the heating process.
From $Q = M\,S\,\Delta T$, the increase in temperature of the water = $2.86 \times 10^7 \div (4210 \times 150) = 45.3°C$.
The final water temperature therefore = 45.3 + 21.8°C = 67.1°C

3. **Q** If the steam produced in Question 1 was passed through unlagged pipes on its passage to the steam-jacketed vessel and the dryness fraction reduced from 1.0 to 0.94, what temperature would the water reach?

A Because some steam has condensed in the unlagged pipework, the latent heat content of the steam has been reduced and now = $0.94 \times 2.86 \times 10^7$ J = 2.69×10^7 J
The increase in temperature of the water therefore = $2.69 \times 10^7 \div (4210 \times 150)$ = 42.6°C and the final temperature = 64.4°C.

4. **Q** The data in the table below represent the different layers through which the latent heat from steam has to be conducted in a stainless steel steam-jacketed vessel when scale and air are present and water is heated. What is the overall heat transfer coefficient of the system?

	Thickness (mm)	K (W/m K)
Air film	0.2	0.03
Condensate film	0.1	0.60
Scale	0.2	1.00
Pan wall	3.0	17.0
Water boundary layer	0.4	0.60

A Using Eqn 38.4:

$$U = \frac{1}{\left[\frac{0.2}{0.03} + \frac{0.1}{0.6} + \frac{0.2}{1.0} + \frac{3.0}{17} + \frac{0.4}{0.6}\right]} \times 10^{-3}$$

(The factor 10^{-3} is required to convert from mm to m.)
$U = 127$ W/m^2 K

5. **Q** What would happen if:
(a) the vessel described in Q4 was made of copper ($K = 386$ W/m^2 K)?
(b) the scale was removed?
(c) the air film was halved?
(d) the air film was removed?
(e) the air film was removed, the scale layer eliminated and the water boundary layer halved?

A
(a) $U = 130$ W/m^2 K
Even though copper conducts heat over 20 times more easily than stainless steel, this alone has little effect on the overall conductivity.
(b) $U = 130$ W/m^2 K
Removing the scale alone similarly has little effect.
(c) $U = 220$ W/m^2 K
Removing just 0.1 mm of air gives rise to an approximate 75% increase in thermal conductivity.
(d) $U = 827$ W/m^2 K
If the air film is completely removed there is a huge increase in U, the value being over six times greater than when air was present.
(e) $U = 1478$ W/m^2 K
Reducing the film layers that offer the largest resistance to heat conduction has the greatest effect on increasing heat transfer.

6. **Q** A steam-jacketed vessel of heated area 2.0 m^2, using steam at 2×10^5 Pa, was found to evaporate 11.2 kg of water in a 5-minute period. Assuming the latent heat of vaporization is 2.20×10^6 J/kg, what is the value of U, the overall heat transfer coefficient (OHTC) of the system?

A The heat energy used to evaporate the water = $11.2 \times 2.2 \times 10^6$ J = 2.464×10^7 J.
The rate of heat transfer occurring is 2.464×10^7 J in 300 s = 8.213×10^4 W (82.13 kW).
From Eqn 38.5: $Q/t = U A\,\Delta T$
The temperature of steam at 2×10^5 Pa = 120.4°C and therefore the temperature difference between the steam and the boiling water, $\Delta T = 120.4 - 100°C = 20.4°C$.
Therefore, $U = 8.213 \times 10^4 \div (2 \times 20.4)$ = 2013 W/m^2 K.

REFERENCES

Arpaci, S.A., Kao S., Selamet, A. (1999) *Introduction to heat transfer*. Prentice Hall, New Jersey.
Long, C.A. (1999) *Essential heat transfer*. Pearson Education, Harlow.
Mills, A.F. (1999) *Basic heat and mass transfer*, 2nd edn. Prentice Hall, New Jersey.