21 Pharmaceutical Biotechnology

Chapter Objectives

At the conclusion of this chapter the student should be able to:

1. Describe pharmaceutical biotechnology and understand the idea behind such therapeutics.
2. Understand the differences between small and large molecule therapeutics.
3. Identify currently available recombinant protein-based pharmaceuticals.
4. Know how characterization leads to the successful development of macromolecular pharmaceuticals.
5. Define physical degradation and give examples.
6. List the common types of chemical degradation reactions.
7. Describe biotechnology product formulation challenges.
8. Discuss the approaches used to produce stable formulations of peptides, proteins, nucleic acids, and viruses.
9. Define the acronym GRAS and understand its importance in formulation.
10. Understand and describe the differences in the ultimate goals of the formulation of biotherapeutics and vaccines.

Introduction

Up to this point, the text has primarily been concerned with drugs of molecular weight under a few thousand ("small molecules"). Nevertheless, much larger molecules, such as proteins, DNA, and carbohydrates as well as macromolecular assemblies including viruses and bacteria, have been used as drugs and vaccines for quite some time and many are currently in development. For example, animal-derived hormones such as insulin and somatotropin (growth hormone) as well as human blood–derived proteins such as coagulation factors and immunoglobulin (antibody) preparations have saved millions of lives during the last century. The availability of synthetic versions of such materials fall into a subclass of pharmaceutical products derived by a general series of procedures known as biotechnology.

Here, in keeping with the general theme of physical pharmacy, the focus will be on what is often called "pharmaceutical biotechnology." There are many comprehensive texts[3,4,5,6,7,8] that discuss the more general aspects of biotechnology with extended discussions of fermentation (the major emphasis of biotechnology until fairly recently), industrial enzymes, and related topics. In this chapter, the analysis, preformulation and formulation of large molecules intended for pharmaceutical use, generally focusing on proteins and nucleic acids, as well as vaccines will be introduced. Methods of their production and delivery will also be briefly discussed but the interested student should refer to more detailed discussions in these areas.[3,4,5,6,7,8]

Key Concept

Defining Pharmaceutical Biotechnology

The Oxford American Dictionary defines this as: "the exploitation of biological processes for industrial and other purposes, esp. the genetic manipulation of microorganisms for the production of antibiotics, hormones, etc."[1] The Oxford Dictionary of Biochemistry and Molecular Biology somewhat more elaborately defines it as: "the integration of natural sciences and engineering sciences in order to achieve the application of organisms, cells, parts thereof and molecular analogues for products and services (European Federation of Biotechnology General Assembly, 1989); a field of technological activity in which biochemical, genetic, microbiological, and engineering techniques are combined for the pursuit of technical and applied aspects of research into biological materials and, in particular, into biological processing. It includes traditional technologies such as fermentation processes, antibiotic production, and sewage treatment, as well as newer ones such as biomolecular engineering and single-cell protein production."[2]

The basic idea in pharmaceutical biotechnology is to employ biological processes and biological molecules to create drugs and vaccines. Our ability to do this has arisen from a dramatic increase in our understanding of the molecular and cellular basis of life. This has included the ability to create and manipulate both nucleic acids and proteins through an extensive series of procedures that is usually referred to as molecular biology. These methods will be very briefly discussed below. One consequence of this dramatic technology expansion has been the rise of the biotechnology industry. This is directly manifested in the creation of a series of companies such as Genentech, Amgen, Genzyme, Biogen-Idec, MedImmune, and many more. In addition, there also exist many hundreds of smaller biotech businesses spread throughout the world focusing on an extensive variety of human diseases employing a diverse array of biotechnology-related technologies. Furthermore, previously small molecule focused large pharmaceutical companies such as Pfizer, Merck, GlaxoSmithKline, Wyeth, and Novartis all contain within

themselves large biotechnology divisions. Currently, there are more than a hundred approved peptide and protein pharmaceuticals (see http://www.drugbank.ca) with hundreds more in clinical trials. It seems fair to say that the clinical importance of large molecules is approaching that of their smaller cousins. The recent mapping of the human genome has further opened up new opportunities with the identification of several tens of thousands of genes providing both new targets and potential new macromolecular drugs. Opportunities for both improved and novel diagnostic procedures have also appeared as a result of the advances mentioned above.

Types of Biotechnology-Derived Products

Peptides and Proteins[9,10,11]

Peptides and proteins are formed by the creation of a peptide bond between combinations of the 20 naturally occurring amino acids. The primary distinction between the two is one of size with polymer lengths less than 30 to 40 residues defining peptides and longer sequences, proteins. A second distinction is one of levels of structure. While both peptides and proteins contain defined orders of their amino acids (their primary structure or sequence), proteins also usually contain additional higher-order structures (Fig. 21-1). For example, most proteins contain regions of regular, local chain interactions known as secondary structure. The two most common types of secondary structure in proteins are the α helix and β sheet. Various types of turns (reversals of chain direction) and more disordered regions are also commonly present. Recently, it has been recognized that some proteins also exist in fairly disordered states, at least in their purified forms. Large peptides may also contain significant regions of secondary structure.

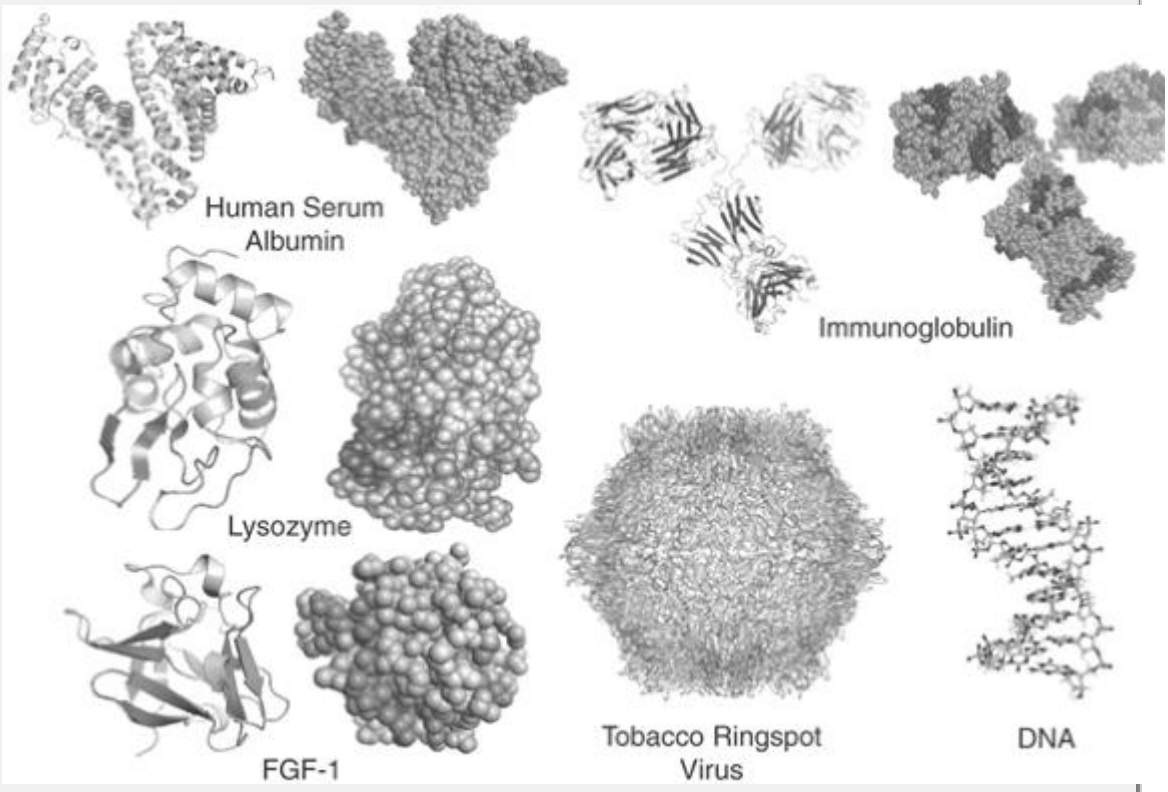

Fig. 21-1. Some typical structures of biopharmaceuticals. Several of these (human serum albumin (HSA), IgG, lysozyme, a fibroblast growth factor (FGF), and DNA) are used to produce data in later figures. Note the highly helical nature of HSA and lysozyme and the extensive β-structure present in IgG and FGF. In the first four cases, the structures are shown in both ribbon and space-filling forms.

Proteins can also bring their various elements of secondary structure together to form what we refer to as tertiary structure. This creates a distinct three-dimensional structure for most proteins. Modern methods of x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy often allow the location in 3-D space of the individual atoms in a protein to be determined with a resolution of 1 to 3 Å. This has resulted in very detailed pictures of thousands of proteins with more appearing each day (Fig. 21-1). One conclusion that has been reached from such work is that distinct, common “domains” (small compact regions of various secondary structure combinations) exist in most proteins, reflecting both evolutionary and functional relationships among many proteins. It is important to realize, however, that the static picture of proteins, seen by crystallography, fails to provide a complete representation of protein structure. Proteins exist in a highly dynamic equilibrium of various conformational states. This will be considered in more detail below.

Individual proteins can also associate into defined multisubunit assemblies forming what is known as quaternary structure. This can involve either multiple copies of the same proteins or heterogeneous mixtures of different types of subunits. There also exists another way in which proteins can associate with themselves, which is referred to as aggregation (Fig. 21-2). This type of structure is especially important to the pharmaceutical scientist since it constitutes a major pathway of physical degradation for many protein pharmaceuticals. This process can be highly ordered (as seen in crystallization or the assembly of fibers) or highly disordered forming amorphous protein particles. We will further consider this latter process below.

Peptides and proteins of pharmaceutical utility can conveniently be placed into a number of different classes. A wide variety of peptide-based drugs are now available. These include antitumor agents such as leuprolide, diabetes drugs

such as insulin and exenatide (Byetta), immunosuppressants like cyclosporine, and labor-inducing agents like oxytocin. Peptides can be either isolated from natural sources such as animals, bacteria, or fungi, or chemically synthesized. They are usually sufficiently small that they can in many ways be treated like lower–molecular-weight pharmaceuticals.

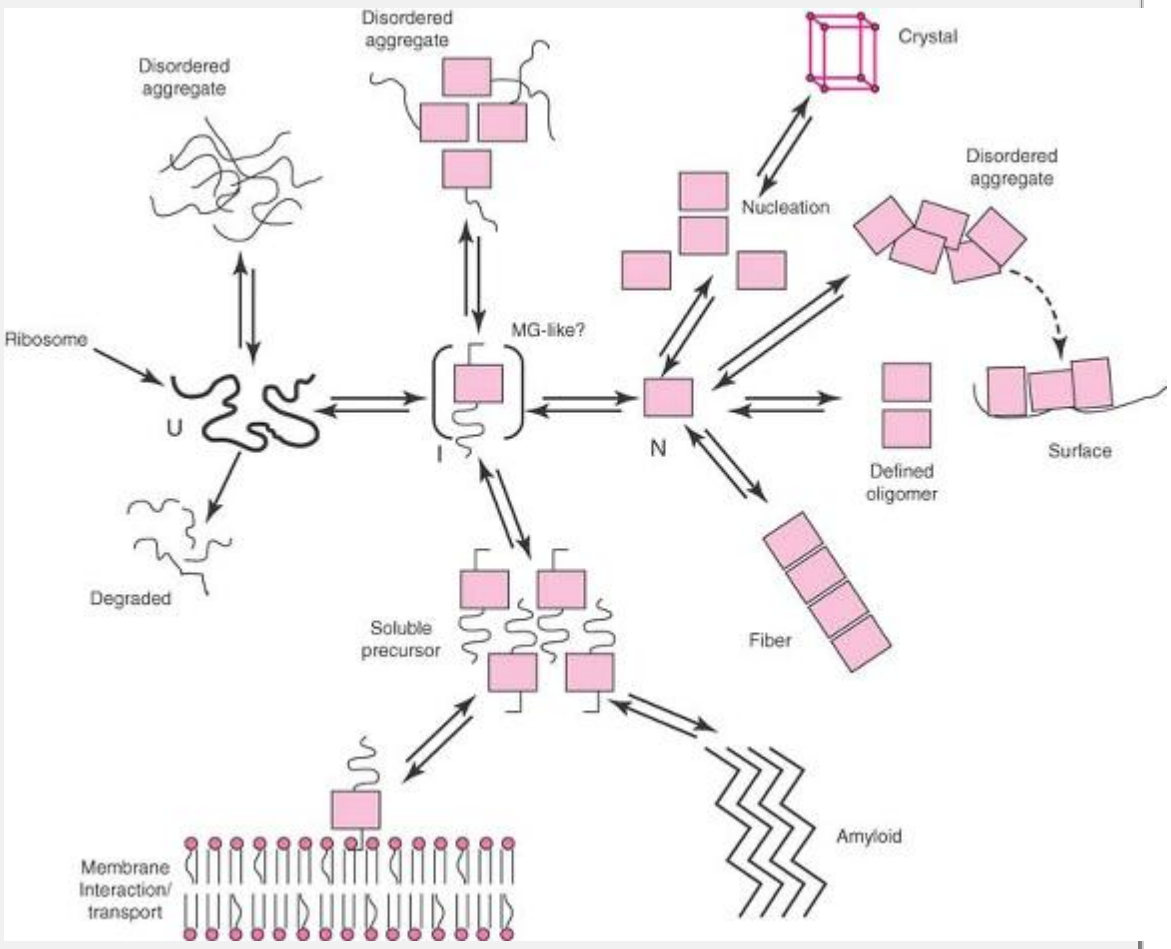

Fig. 21-2. Aggregation of proteins is a major pathway of their physical degradation. The native state of a protein can associate into ordered species that are crystalline or fibrous in nature and assemble into defined oligomeric species such as dimers and tetramers or into amorphous aggregates. Under various forms of stress, some structure can be lost, often into forms known as molten globule states, which can also form disordered aggregates or associate into soluble aggregates which form amyloid materials or become surface active and bind to membrane or container surfaces. Native assemblies can also frequently be surface active. These intermediate states are the most common origin of aggregation problems in protein pharmaceuticals. More complete unfolding is rarely encountered because these intermediate states dominate but unfolded (denatured) protein can aggregate as well. (Modified from C. M. Dobson and M. Karplus. The fundamentals of protein folding: bringing together theory and

experiment. Curr. Opin. Struct. Biol. **9**, 92–101, 1999.)

Proteins that are used as drugs currently cover such a wide range of diseases that it is difficult to summarize their activities and applications. Only a few examples are provided here. Many of the early pharmaceutical proteins were derived from human or animal blood or tissue. Insulin and somatotropin (growth hormone) were originally derived from animal pancreases and human brains, respectively. Serum albumin, coagulation factor XIII, and the hepatitis B surface antigen were all obtained from human blood. Various antivenoms against snake and spider venoms were (and still are in many cases) obtained from the blood of large animals like horses and goats. Most of these are now, however, obtained recombinantly, reducing or eliminating problems of immunological reactions, contamination, expense, and supply.

Many other proteins of potential therapeutic use were initially impossible to obtain from natural sources in sufficient amounts or quality (purity) to make them realistic candidates for use as therapeutic agents. Many such proteins are now produced recombinantly. While in some cases still expensive, the recombinant forms are both safer and available in abundant supply.

Currently Available Recombinant Protein-Based Pharmaceuticals

Currently available protein-based drugs can be subdivided into several categories. These are (based on 2008 sales)

P.519

hematopoietics (23%), monoclonal antibodies (20%), cytokines (19%), vaccines (13%), antithrombins (11%), plasma proteins (6%), insulin (5%), and growth hormones (2%). (Information cited in genengnews.com, January 1, 2009.) A brief discussion of a few representative examples follows.

Tissue plasminogen activator (t-PA) is a protease, which can dissolve thrombi which form at sites of coronary vessel occlusion and can induce myocardial infarction and stroke. Removal of such blockages within the first few hours of a coronary event can be lifesaving. Previous therapy using proteins such as urokinase and streptokinase was less specific, often leading to internal bleeding problems. Because t-PA binds to fibrin and is a natural human protein, it is more specific with generally less side effects. The protein must be produced in mammalian cell culture because it is glycosylated in its natural form (see below). Erythropoietin (EPO) is another example of a protein drug with a dramatic pharmacological effect. This protein stimulates the production of red blood cells (erythrocytes) and has a variety of uses including treatment of kidney dialysis patients to prevent anemia. Like t-PA, the protein is produced in mammalian cells because it must be correctly glycosylated for full biological activity. Similarly, granulocyte macrophage colony stimulating factor (GM-CSF) and granulocyte colony stimulating factor (G-CSF) are used to enhance the proliferation of hematopoietic progenitor cells important for immune responses and are used during bone marrow reconstitution as well as some cancers.

A variety of proteins known as interferons and cytokines are now available in recombinant form and are used for a wide variety of disorders. Interferon-β is used to effectively treat multiple sclerosis (MS) and α-interferon as an antiviral agent. The interleukins also belong to the cytokine family. The best known of these is interleukin-2 and is most commonly used for the treatment of renal cell cancers. A wide variety of interleukins (more than 20 are currently known) are being explored for immune-related applications. Some other notable therapeutic proteins include the Factor VIII and IX coagulation factors, DNAse for the treatment of cystic fibrosis, and glucocerebrosidase for enzyme replacement therapy in Gaucher disease. These examples are fairly typical in that they either target receptors or their ligand or are themselves ligands or receptors or fragments or mutants, thereof. Alternatively, they are proteins with enzymatic activities of therapeutic consequences. Recently, however, the biotechnology industry has begun to be dominated by the class of proteins known as immunoglobulins (Igs). When these proteins have defined affinity for specific ligands they are known as antibodies. All Igs consist of two heavy chains and two light chains bound covalently by disulfide bonds (Fig. 21-1). The chains themselves are

composed of two globular domains in the case of light chains and four or five in the case of the heavy chains. Each domain comprises a beta-sandwich-like structure with the two flat sides of the sandwich also held together by disulfide bonds. The five different classes of immunoglobulins (IgG, IgA, IgD, IgE, and IgM) are defined by differences in their heavy chains. There also exist two different types of light chains known as lambda and kappa. The N-terminal part of the heavy and light chains varies significantly in three regions of their sequence known as hypervariable regions, a phenomenon which arises because of controlled genetic recombination events at the DNA level combined with somatic mutation. These highly variable regions are brought together in space to form millions of different binding clefts sufficient to recognize with both low and high affinity most substances, which are referred to as antigens. This generates a huge library of receptorlike molecules that can be used to create pharmaceutical proteins that can interact with virtually any chosen target. Not surprisingly, this has resulted in the use of Ig as therapeutic agents for amazingly diverse applications. Of the 5 Ig classes, it is the IgG type that is generally used. IgG itself consists of several subclasses with differing biological properties. By using cellular cloning methods, it is possible to create unique (monoclonal) recombinant antibodies as well as their fragments for virtually any ligand. Although the original technology was developed for mouse antibodies, it is now routinely possible to produce entirely human antibodies or animal antibodies in which the nonvariable parts of the antibodies are converted to human form (humanized monoclonal antibodies, hmAbs). Although IgG molecules are large (~150 kDa) and glycosylated (usually necessitating their production in mammalian cell culture), their unique specificity and long serum half-life contribute to their expanding use as therapeutic agents. A list of some currently marketed antibodies including their target and use is shown inTable 21-1. This diversity of applications ensures their continued development as therapeutic agents well into the immediate future.

*Vaccines*12,13

Outside of public health measures, there is little doubt that vaccines have had the greatest positive effect on human health. Although vaccines have now been used for several hundred years, there has been a resurgence of interest in their use in the last decade due to the new technologies available for their creation and an improvement in their financial viability.14 Vaccines function by exposing our immune systems to attenuated pathogens, pieces of pathogens, or other agents (all referred to as “antigens”). Under the right conditions (still incompletely understood), this can produce a “memory” response which results in a very robust immune response that can protect the immunized individual against later exposure to actual disease causing pathogens. There are three divisions of the human immune response generally recognized15,16 (Fig. 21-3). The first is the innate response which primarily involves the recognition of repetitive structures on the surface of pathogens by receptors on immune cells known as “toll” receptors. This leads to a complex series of cellular responses including activation of the adaptive immune response involving the production of antibodies (the humoral response) and the cellular response

which among other activities produces cells which can kill infected host cells. When testing vaccines, the production of antibodies is the most common event measured although it is becoming common to quantitate the production of cytokines as an indicator of cellular responses.

Table 21-1 Currently Marketed Antibodies Target and Use

Name	Target	Typical Uses
Avastin	VEGF	Cancer (multiple)
Bexxar	CD20	Non–Hodgkin lymphoma
Campath	CD52	Leukemia
Erbitux	EGFR	Colorectal cancer
Hercetin	HER-2	Breast cancer
Humira	TNF-alpha	Arthritis
Mylotarg	CD33	Leukemia
Orthoclone OKT3	CD3	Prevent transplant rejection
Raptiva	CD11alpha	Psoriasis
Remicade	TNF-alpha	Immune inflammatory disorders
Reopro	GPIIb IIIa	Inhibits platelet aggregation (thrombux formation)
Rituxan	CD20	Non–Hodgkin lymphoma/arthritis
Simulect	CD25	Inhibits allograft rejection
Synagis	RSV	Treats RSV infections
Tysabri	VLA4	MS/Crohn disease
Xolair	Ig E	Allergic asthma

Zenapax	IL-2	Inhibits allograft rejection
Zevalin	CD20	Non–Hodgkin lymphoma

A wide variety of different antigens have been employed as vaccines.12 Among the most effective are those that employ attenuated viruses. Vaccines such as measles, mumps, rubella, varicella, rotavirus, and one form of the polio vaccine are all examples of such vaccines and have had a dramatic impact on human health. Effective vaccines can also be created by inactivating viruses and bacteria through chemical or radiation methods, with the hepatitis A, rabies, and some forms of the polio and influenza vaccines representative examples. From early in the history of vaccinology there has been the hope that individual components of organisms could also be used as the active components of vaccines. Today, proteins, carbohydrates, and nucleic acids are all important antigens with the former two currently employed in marketed vaccines. Originally, proteins for vaccine use were purified from actual pathogens. These include inactivated proteins such as typhoid and cholera toxins, partially purified influenza proteins (primarily the hemagglutinin and neuraminadase), as well as viruslike particles from the serum of hepatitis B infected individuals. More recently, recombinant methods (see below) have been employed, although there has been only limited success in these efforts. Nevertheless, the hepatitis B surface antigen (HBV), cholera toxin B, and surface proteins from the human papillomavirus (HPV) are all successfully used in highly effective vaccines. The OspA protein from the organism causing Lyme disease was also employed but the vaccine was eventually removed from the market. It should be noted that two of these recombinant vaccines (HBV and HPV) are in the form of viruslike particles which significantly enhance their immunogenicity. Several vaccines that employ carbohydrates as antigens are also available. In the case of adults, a vaccine against pneumonia has been developed which contains 23 different purified polysaccharide chains. Polysaccharides are only weakly immunogenic in children, however, and must be conjugated to protein-based carriers for them to be effective in infants. These carriers include diphtheria and tetanus toxoids as well

as an outer membrane protein complex from *Neisseria meningitides*. Important childhood vaccines based on polysaccharide/protein complexes include those for *Haemophilus influenzae*, meningoccal disease, and pneumococcus. Finally, it has recently become apparent that it is possible to make vaccines using DNA plasmids in which antigenic proteins are encoded within appropriate expression sequences. Through mechanisms that are still not entirely understood, expression of these antigenic proteins by cells of the immune system or their secretion by other cells can lead to significant cellular and humoral immune responses. Although successful human vaccines have yet to be developed on the basis of this technology, two veterinary vaccines are on the market. In another emerging technology, the integration of specific protein sequences into virus delivery vehicles such as adenoviruses is also being explored. Perhaps the most important points to note here are that (a) there are a wide variety of different

approaches to vaccines and (b) vaccines themselves are usually quite complex presenting significant challenges for their formulation and delivery.

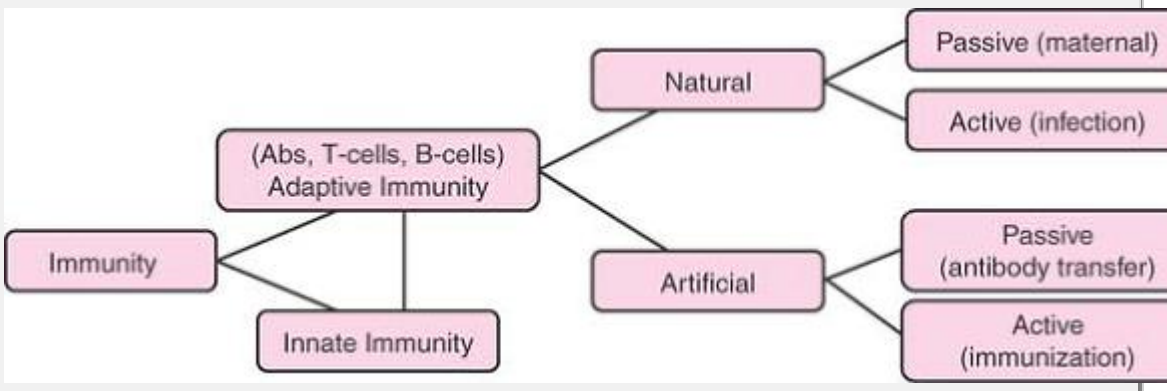

Fig. 21-3. A basic diagram of the mammalian immune system (see text).

Nucleic Acids[17,18,19,20]

As indicated in the previous section by the example of DNA vaccines, nucleic acids in both their DNA and RNA forms can also be used as pharmaceuticals. Historically, an approach known as antisense RNA was used to either alter gene expression or interfere with the translation of RNA into protein. The idea is simple in principle but difficult in execution. A piece of single stranded RNA that has a sequence complementary to a gene or RNA of interest (the "sense" target) is introduced into appropriate cells. The highly specific binding of the antisense drug to a specific mRNA or gene may result in its destruction through an enzymatic process or blocking of transcription, respectively. Various chemical analogues of RNA are often used for this purpose to improve their stability. The delivery of antisense RNA pharmaceuticals into cells, however, has presented a formidable challenge which remains unsolved despite a wide range of attempts to overcome the barriers that exist to this problem. So far, only a single antisense drug has reached market (for treatment of cytomegalovirus retinitis). Currently, the enthusiasm for this approach is rather low with approaches based on RNA interference (RNAi) (see below) more actively pursued.

A second approach relies on the introduction of a gene (DNA) coding for a protein of potential therapeutic activity into cells (i.e., gene therapy). This gene (or genes) is usually introduced either incorporated into the genome of a virus21 or as a part of a bacterial plasmid, the latter often complexed to a polymeric, (usually) cationic carrier to facilitate entry into cells.[22,23] Initially, RNA retroviruses with the ability to integrate their genes into host chromosomes were the most actively employed viral vectors. Problems of toxicity, however, have reduced their use. More recently, the DNA viruses, adenovirus and adeno-associated virus have been used. For safety reasons, there is also significant interest in using bacterial plasmids to delivery genes for therapeutic applications. Plasmids are circular, double-stranded, facilitated self-replicating pieces of DNA that can contain multiple genes as well as auxiliary sequences that can aid in replication and gene expression. They are easily produced in high numbers in host bacteria or other cell types making their manipulation and manufacture at high concentration easier than viruses. In most cases, plasmids are complexed to cationic polymers such as positively changed lipids or polyethyleneimines to facilitate cellular entry and to enhance their stabilization. The use of plasmids provides an especially flexible approach to gene therapy given their large genomic capacity and ease of synthesis, but so far no human therapeutics have been directly derived from this technology. Again, problems with their delivery, potency, and safety have inhibited their development.

The most recent approach to the use of nucleic acids as therapeutics has come from the discovery that many genes are naturally regulated by the presence of small RNA molecules.24,25This is a rapidly evolving field with new discoveries being made almost weekly. From a pharmaceutical perspective, this new technology clearly holds significant promise. Initially, efforts have been focused on the use of 21 to 23 nucleotide double-stranded RNA molecules which are complementary to target mRNA species. In a manner similar to but distinctly different from antisense activity, the complexes formed are subject to destruction by a naturally occurring catalytic activity. The biochemistry and cell biology of RNA interference is both complex and fascinating.19 We leave it to the interested student to pursue this topic further on their own.20Promising effects have been seen in animal disease models and human clinical trials are underway.24,25

The potential use of “aptamers” as therapeutic agents will be briefly mentioned. These are small DNA or RNA molecules (typically 15–60 nucleotides in length) that have been specifically selected for their specificity and affinity for proteins or other biological targets.26,27,28In fact, a drug for age-related macular degeneration is already available based on the use of aptamers. A number of clinical trails are also underway employing these unique molecules. Peptide-based aptamers are also under investigation. These generally consist of a variable peptide loop that is incorporated into a protein matrix.

Discovery of Biotech Drugs

Small molecule drugs are typically discovered by screening libraries of natural or synthetic compounds (combined with rational structure optimization) against protein-based targets. Targets are selected on the basis of our current fairly extensive understanding of metabolic and hormonal pathways and extrapolated therapeutic effects. It is this same modern understanding of cell and molecular biology29,30accompanied by corresponding pathologies that is used to design biotechnology-based drugs. One difference, however, is that in many cases components of these pathways or their

analogues are used as actual drug substances. Thus, protein drugs such as insulin, human growth hormone, and EPO are simply used to supplement their naturally occurring counterparts. With the availability of the sequence of the human genome of approximately 20,000 to 30,000 genes (the exact number remains quite controversial), in principle at least, all such genes are now available for therapeutic use. In addition, the many variants of each gene (sometimes in the hundreds and known as single-nucleotide polymorphisms or SNPs) are being increasingly defined and offer the potential for more specific therapeutic use. In addition, the use of antisense RNA and more recently RNA interference offers the possibility of gaining fairly detailed functional information. This can often be used to create cellular models that provide useful analogues of specific biochemical pathways or even disease states that can be employed to further develop protein and nucleic acid therapeutics. In ideal cases, animal models either natural or transgenic in nature in which specific genetic changes have been made to simulate human diseases can be used. This subject is a vast one and rapidly changing, but it seems certain that the information necessary to develop recombinant pharmaceutics will become increasingly available.

*Cloning*31,32

Approximately 30 years ago, Paul Berg, Herbert Boyer, and their colleagues realized that it should be possible to manipulate DNA in such a way that specific genes could be inserted in cellular systems and their expression induced. This has led to a now routine series of procedures to accomplish this task. Although the details of these methods can be quite complex and an art form in themselves, the basic idea is straightforward.31,32 First, a specific gene (e.g., one that is selected as a potential protein pharmaceutical) must be isolated. This is typically done by screening a large library of genes that have been inserted into circular pieces of bacterial DNA (plasmids). This gene is then inserted into a plasmid which has been specifically designed for expression of the gene into protein (an “expression vector”). To accomplish this, the gene is removed from the library plasmid by cutting with a highly specific protein known as a “restriction enzyme.” The expression plasmid is then opened with the same enzyme and

combined with the desired gene. The gene hybridizes to the sticky ends of the plasmid in a highly specific manner. The host plasmid is then covalently closed with another enzyme known as a ligase. The “recombinant” plasmid can then be inserted into a host cell for reproduction (we say the cell is transformed, a process not to be confused with the transformation of cancer cells).

A wide variety of different cells types are available for expression purposes. These include bacteria such as *Escherichia coli*, yeast cells, baculovirus, animal cells in culture, plants such as corn and tobacco, and transgenic animals like goats, sheep, and cattle. It is also possible to express proteins in cell-free systems using extracts of mammalian cell cytoplasm. There are advantages and disadvantages to each expression system. For example, low cost and high levels of expression often dictate the use of bacterial and yeast cells. If posttranslation modifications such as glycosylation are necessary for the proper functioning of the protein, then eukaryotic, yeast, or baculovirus systems are typically used although each may produce a uniquely modified protein. If larger proteins are being expressed, then eukaryotic cells are usually employed.

The vectors used to transform the expression system of choice must meet a number of requirements. They must contain one or more appropriate promoter sequences (binding sites for RNA polymerase), an origin of replication for DNA polymerase, a ribosomal binding site, appropriate restriction sites, and termination sequences. In addition, selectable genes that when expressed allow cellular survival when stressed and affinity tags to aid in the isolation of the expressed protein are often included (see Fig. 21-4).

Common promoters include those from phages, viruses, and inducible systems such as the arabinose system. A wide variety of tags are available for purification purposes. The most common is the His6 sequence, which can be used for Nickel affinity chromatography, and the FLAG epitope sequence of DYKDDDDK. Others include the cellulose-binding domain and glutathione-S-transferase tags as well as the c-myc epitope (EQKLISEEDL), poly E 35 and poly K35, and CBP-calmodulin-binding domain tags.

The next step is to get the DNA into the host cell (transformation). In the case of bacteria, this is usually accomplished by making the plasmid particulate through the addition of a positively charged ion such as calcium, followed by rapidly lowering the temperature to damage the bacterial membrane, facilitating plasmid entry. A more modern and efficient method is through the use of a process known as electroporation. This uses a pulse of electromagnetic energy to open up pores in the bacterial membrane. Although this procedure is

still poorly understood, it works for many types of cells and is even being considered for human use and gene therapy. Genes can also be introduced into cells by using viruses as cloning vectors (which have natural receptors) or by complexing the recombinant plasmid to a positively charged agent such as a cationic lipid or other positively charged polymer. To some extent, the process is an empirical one with the extent and stability of expression the criteria for success.

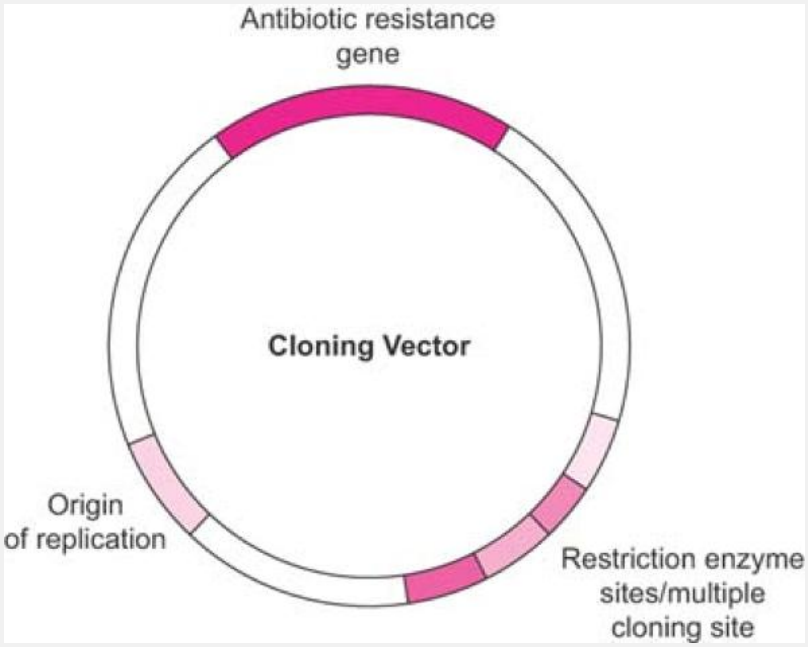

Fig. 21-4. The basic design of an expression plasmid. The actual plasmids used in the productions of biopharmaceuticals are usually much more complex (see references 31 and 32).

Transformation of yeast and mammalian cell lines ("transfection") have some similarities to bacterial systems but some key differences as well. The student is referred to the more technical literature to explore these differences, which include selection procedures, plasmid design, and cellular growth requirements among others.[3,4,5,6,7,8,31,32]

Once cellular transformation has been performed, it is necessary to grow these cells to a high mass to produce sufficient material for isolation of the target product (e.g., protein, virus, DNA plasmid). This is accomplished by a process known as fermentation. This procedure varies depending on the cell type, but in general, variables such as CO_2 requirement, O_2 levels, and necessary nutrients must be individually optimized. Typically, microbial cells grow much more rapidly than animal cells and grow to higher cell densities. Growth is usually performed in large fermentors or bioreactors, which permit the growth process to be continuously monitored and the amounts of O_2, CO_2, and nutrients to be maintained at appropriate levels. The use of incubators and fermentors is a highly specialized activity and requires extensive knowledge and training for maximum effectiveness.

Purification of Macromolecular Therapeutics[33,34]

After sufficient growth of cells and subsequent expression of the target protein or other macromolecular drug agent, the product can be found either inside the cells, secreted into the surrounding medium, or perhaps associated with the cells' surface. If the protein (or DNA, or virus) is secreted into the growth medium, its purification can usually begin immediately. If it is found inside the cell, however, its isolation is usually much more difficult. The cells must initially be broken open. This is usually done by using a French pressure cell (shear), sonication (sound), or disruption with glass beads (mechanical stress). All three methods are relatively rough, so gentler methods are also often used. These include freeze-thaw stress, lysis with detergents (dissolution of cell membranes), and enzymatic or osmotic lysis. Once the macromolecular drug substance is released from its association with the transformed cells, it must then be isolated from the contaminating proteins, lipids, carbohydrates, and nucleic acids. This is usually a relatively complex multistep process that is primarily empirically based although the physical properties of the protein/nucleic acid/virus can often provide important clues to the most effective steps.

Initial steps in the purification of macromolecular substances typically involve crude separations based on the gross solution behavior of the material. For example, differential precipitation by salts, organic solvents or solutes, pH, or temperature is frequently used for this purpose. In some cases, this may be followed by filtration steps using filters containing micron or submicron-sized pores that selectively pass or retain the macromolecular drug.

The primary class of higher resolution approaches used today involves a variety of types of chromatography.[35,36] Most generally, these procedures rely on passing a mixture of molecules in a mobile (liquid) phase, through a stationary (solid) phase, resulting in a partial (usually) or complete separation of the components. This occurs because of an interaction or lack thereof between the molecules to be separated and the solid phase. There are currently many different types of chromatography, some used analytically (to be discussed below) and some to separate molecules at preparative levels. The four most widely used approaches for the latter purpose are described below (Fig. 21-5).

If it can be successfully performed, some version of “affinity” chromatography is usually an optimal choice. This is

because of the degree of purification (often greater than a thousandfold) frequently obtained by this method. The basic principle behind affinity chromatography is that highly specific interactions often exist between proteins and protein complexes like viruses and other molecules (referred to as “ligands”). If a ligand can be fixed (usually covalently) to an insoluble matrix (often in the form of small beads), then this can be used as a basis for separation since most or all of the other components in the mixture will not undergo this specific interaction. It is, of course, necessary to remove the molecule of interest from the support, but this can often be achieved by some type of competitive interaction with another substance or alteration of the solution conditions by a variable such as (low) pH or (high) ionic strength. The trick is not to alter the desired macromolecule by the elution process. Ligands are typically attached covalently to a resin such as sepharose while retaining an affinity in the range of K_D 10^{-4} - 10^{-8} M. A wide variety of ligands have been used for protein affinity chromatography. For example, proteins with polyhistidine tags will bind to columns containing immobilized metal ions such as Zn^{+2} and Cu^{+2}. They can generally be eluted with either imidazole which competes for the binding or low pH (e.g., ~4.5). The problem with this approach is that the protein drug now contains the His-tag, often an undesirable modification of its native structure (His-tags can be “sticky” because of their positive charge at pH <5). This problem can be overcome by using a cleavable linker between the proteins and the tag, but in general, His-tag affinity chromatography is primarily used in the discovery and basic research stages of pharmaceutical development. In contrast, a method that is commonly used to purify large amounts of immunoglobins uses protein A and G. These proteins bind to Igs with high affinity. They have been attached to various resins and are routinely used in immunoglobulin purification with elution performed by lowering the pH. Another commonly used ligand is heparin or other highly polyanionic polymer. Many proteins contain high affinity (although not necessarily highly specific) polycationic sites that can interact electrostatically with polyanions.[37,38] For example, many growth factors, cytokines and blood proteins, contain such sites. High ionic strength is often sufficient to elute such proteins from heparin columns. Other ligands that have been used to isolate proteins of pharmaceutical interest include calmodulin (kinases, phosphatases, cyclases), adenosine monophosphate (AMP) (adenosine 5′-triphosphate (ATP)-dependent proteins, nicotinamide-adenine dinucleotide (NAD) cofactor utilizing proteins), and cibacron Blue dye (proteins with large hydrophobic binding sites such as the serum albumins).

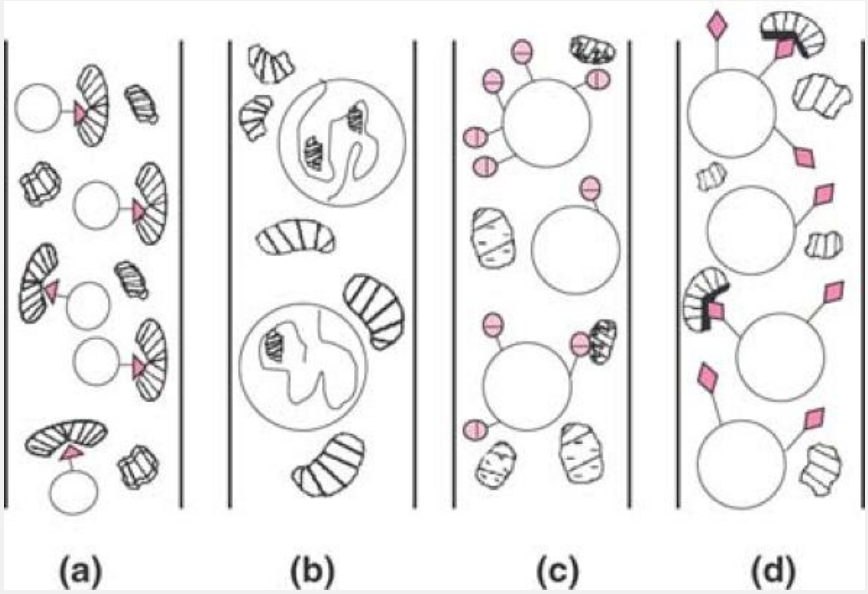

Fig. 21-5. The basic forms of liquid chromatography. The most common forms of chromatography used to purify and characterize biomolecules and their complexes are based on interactions between the biopharmaceuticals and a solid support (indicated by large open circles) which are usually derivatized with a molecular moiety to induce interactions with target molecules. (*a*) In affinity chromatography the immobilized ligand (▲) binds to a specific site on the macromolecule (all shown with lined interiors) and is usually eluted with a related soluble ligand or a change in pH. (*b*) In size exclusion chromatography, no ligand is used but smaller molecules are able to diffuse into particle support interiors slowing their progress through the column resulting in separations based on size. (*c*) In ion-exchange chromatography, macromolecules and their complexes are separated on the basis of charge/charge interactions with the support with elution induced by either changing the pH or increasing salt concentrations. (*d*) Hydrophobic interaction chromatography separates target macromolecules based on the interactions between apolar (hydrophobic) ligands and apolar sites on proteins. Elution is usually performed by lowering salt concentrations (which weakens apolar interactions). The same principle is used in reversed-phase HPLC but elution is typically induced by increasing concentration of an apolar solvent.

A chromatographic method which separates proteins on an entirely different physical basis is gel filtration (also commonly known as size exclusion chromatography or SEC). In this case, proteins, nucleic acids, and even viruses can be separated on the basis of size. The method employs small beads of various porosities. Smaller proteins can enter into the highly channeled interior of these matrices and consequently their progression through a column of such material is slowed. In contrast, larger proteins can less efficiently enter into the beads or pass between them entirely. The overall result is a time-dependent separation of mixtures of proteins, nucleic acids, or even viruses from one another based on their hydrodynamic size. As discussed below, this method can also be used as an analytical tool to estimate the size and molecular weight of macromolecules and to detect aggregated material.
A very commonly used method to separate macromolecules and especially proteins is ion-exchange chromatography. This technique separates molecules on the basis of their charge rather than their size or specific affinity for ligands. To perform this method, either positively or negatively charged groups are covalently attached to a polymer, which permits the free flow of macromolecules. The most commonly

employed groups are the negatively charged carboxymethyl or sulphopropyl moieties and the positively charged quaternary aminoethyl and diethylaminoethyl side chains. The negatively and positively charged groups are referred to as cation and anion exchangers, respectively. Once adsorbed to the charged surface, the macromolecule is eluted by either increasing the ionic strength with salt (in either a step or gradient fashion) or altering the pH to minimize the electrostatic attraction. Ion-exchange resins are classified as either weak or strong. Weak exchangers are typically effective over a more limited pH range than the strong exchangers. The selection of an appropriate ion exchange resin is determined by the macromolecules isoelectric point (the pH at which it has zero net charge), size, pH stability, and the scale at which the separation is conducted.

The fourth method is known as hydrophobic interaction chromatography. In this case, the group which is derivatized to the support resin is hydrophobic (more correctly “apolar”) rather than charged or a specific ligand. Apolar sites on molecular surfaces can bind to these apolar groups. To increase the strength of this interaction, the molecules are usually loaded at high salt concentrations. In contrast to electrostatic interactions, which are weakened at higher salt concentration, hydrophobic interactions are increased under these conditions. Thus, macromolecules are eluted by lowering the salt concentration. This technique bears a resemblance to high-performance liquid chromatography (HPLC), which will be discussed below as an analytical technique. Both employ resin-attached apolar side chains of length C1-C18. HPLC, however, usually elutes protein with an organic solvent like acetylnitrile or propanol. Unfortunately, this usually alters macromolecular 3-D structure (although often, reversibly), often negating its use as an isolation technique. The exception is peptides, which may not require a defined 3-D structure for their biological activity.

Characterization[39,40]

A key to the successful development of a macromolecular pharmaceutical is to obtain a detailed picture of its structure

and the response of this structure to environmental stresses such as temperature, pH, ionic strength, agitation, and freeze/thaw exposure. Although a detailed structural picture of a molecule like that obtained by x-ray crystallography or NMR spectroscopy can be very helpful, this is not absolutely necessary for the successful formulation and development of a biopharmaceutical. The usual approach is to use a variety of different lower resolution experimental approaches to obtain pictures of the molecule from a wide range of perspectives. Some of the most common methods are discussed below.

Biology-Based Assays (Bioassays)[41]

Although often lacking a high degree of precision and accuracy, assays based on the response of animals or cells to biotherapeutics are generally considered to be the ultimate arbiter of the retention of pharmaceutical activity. Well-known examples of the use of animals for this purpose include the lowering of blood glucose in test animals by insulin, an increase in weight upon the injection of growth hormone, and the production of specific antibodies when animals are exposed to vaccines. In some cases, transgenic animals have been developed as disease models.

Thus, diabetic mice, mice with Alzheimer-like disease, and many animals with genetic defects similar to those found in humans have all been created and provide a basis with which to check the effectiveness of a particular biotherapeutic. The advantages of such approaches are obvious. They permit a direct evaluation of the critical properties of the pharmaceutical and may be sensitive to changes in its structure. It is possible, however, for structural changes to occur that are not detected by such methods because either the structural alteration does not affect activity or the usually fairly wide experimental variability in such measurements does not permit detection of relevant structural changes, even if the biological activity is perturbed. For these reasons (among others), additional assays based upon a variety of different physical and chemical properties of the target biopharmaceutical are usually also employed.

Assays based on the response of cells to the presence of drugs are being increasingly used to check for structural integrity and biological activity. Many if not most biopharmaceuticals act on one or more cell

types by either binding to receptors on their surface or entering cells and producing a consequent molecular response. These cellular responses may involve changes in the level of important cellular messenger molecules such as cyclic AMP, ATP, and inositol phosphates or alterations in cellular properties like membrane potential, rate of cell division, or even cell death (toxicity). This is a large subject that we cannot explore further here, but an increasing number of experimental approaches for such measurements are becoming available facilitating this approach. A question that has yet to be unambiguously answered is to what extent cellular-based assays can replace animal studies. The speed, simplicity, and precision of the former, however, point to their increased usage.

Immunoassays[42,43]

Immunoassays are methods that employ antibodies to detect the amount of an antigen (e.g., protein, DNA, polysaccharides). These are solution or solid state assays that can employ either monoclonal antibodies obtained by a variety of methods or antisera obtained from the blood of animals injected with the macromolecule or macromolecular complex of interest. Monoclonals have the advantage that they are highly specific for a single site but antisera can be more easily obtained and their multispecificity can be advantageous under certain circumstances. In the case of vaccines, immunoassays can be used to measure antibodies produced in response to the vaccine. Alternatively, the presence and amount of a biopharmaceutical can be determined. In both cases, either the antibody or the antigen must be labeled in a way that the amount of that component can be easily quantitated. Common forms of labeling include attaching a fluorescent or colored group, a radioisotope (a radioimmunoassay or RIA), an enzyme (an enzyme-liked immunoassay or ELISA), or a magnetic particle (a magnetic immunoassay or MIA). A number of other approaches are also frequently used. For example, addition of antibody to antigen often produces insoluble complexes due to the multivalent nature of the antibody. This can be measured by light scattering or turbidity (see below). The latter is often referred to as nephelometry. The antibody or antigen can also be attached to cells (usually red blood cells) and the aggregation of the cells measured by a method known as agglutination.

Many immunoassays are conducted in a competitive manner. Either a labeled antibody or antigen is used and the unlabeled molecule or virus of interest is used to compete or displace the labeled component in the antigen–antibody interaction (Fig. 21-6).

Electrophoresis[44,45]

The fundamental theory of electrophoresis is described earlier in the book and should be reviewed prior to reading this section. Here the focus is on the versions of electrophoresis that are primarily used for biomolecules, gel, and capillary electrophoresis. Gel electrophoresis is today usually performed in thin gels of cross-linked polymers. When the method is used for proteins, the gel usually consists of cross-linked polyacrylamide, often in gradient form. When nucleic acids such as DNA are being electrophoresed, this is usually performed in agarose. In both cases, a semisolid medium of porous properties is used and an electromagnetic field is employed to move the macromolecule through the matrix. The relative “mobility” of the macromolecule is determined by a number of factors including its size and charge as well as in some cases, its 3-D structure. The most common form of protein electrophoresis is known as sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (SDS-PAGE). In this case, the presence of SDS, a negatively charged detergent, causes proteins to unfold into rod-shaped structures1

Because the SDS binds relatively uniformly to the protein, the rodlike complexes migrate in proportion to their molecular weight (Fig. 21-7). If protein molecular weight standards are included in lanes of the gel not containing the protein of interest, a fairly accurate estimate of the latter's molecular weight can be made from a plot of the logarithm of the molecular weight of the standards versus their position on the gel and comparison of the unknown's migration behavior. If a reducing agent such as β-mercaptoethanol or dithiothreitol is present, any disulfide bonds present will be broken and individual bands for any subunit present will be seen. For example, when immunoglobulins are subjected to SDS-PAGE in the presence of a reducing agent, bands for both the heavy and light chains can be clearly

distinguished and their molecular weights estimated (Fig. 21-7). If no SDS is used, then it is much more difficult to estimate molecular weights but relative size information about proteins can still be obtained, including separation of proteins in mixtures. Because proteins do not usually absorb visible light (with the exception of those containing chromophores such as the heme groups of hemoglobin and the cytochromes), the gels are usually stained with a visible dye. Alternatively, gels can be stained with silver nitrate, a method approximately 50× more sensitive than the most commonly used dye, coomassie blue. The amount of protein present in each band is roughly proportional to the amount of associated dye. This can be quantitated by scanning the gel with a laser or by using some other type of imaging technique. There are a large variety of other types of electrophoresis that can also be used with proteins. For example, the separated protein can be blotted onto another piece of material and this stained with labeled antibodies to identify specific proteins of interest (a “western” blot). This permits a highly reliable identification of proteins to be made. Gradients of a protein unfolding agent like urea or a gradient of temperature can also be imposed. Such techniques permit protein unfolding and subunit dissociation to be studied.

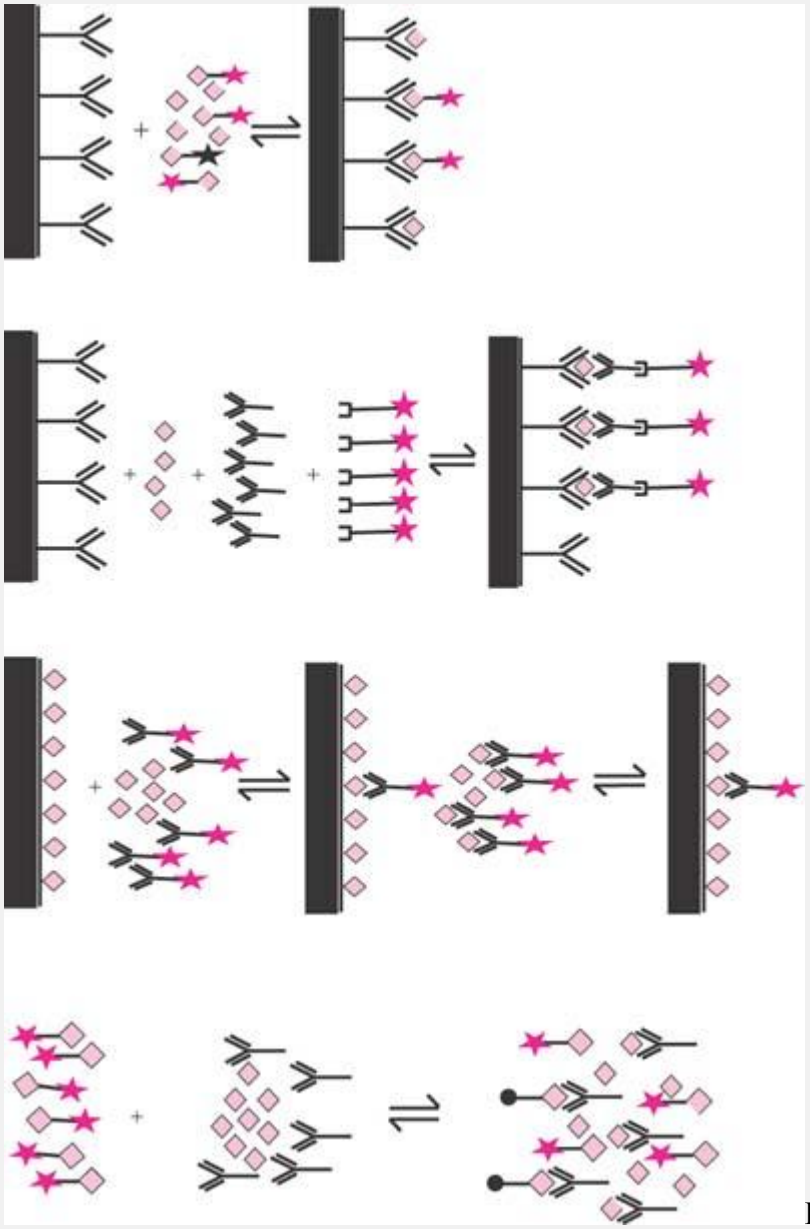

Fig. 21-6. Types of immunoassays, the symbols represent antibodies (⟩–), antigens (♦) labels (•), and a solid support such as the well of a microtiter plate (—). (*a*) A heterogeneous, simple competitive assay in which labeled and unlabeled antigen compete with one another. (*b*) A sandwich assay in which an excess of a labeled reagent antibody competes for a site on the antigen which is captured on the support by another antibody which binds to a different site. (*c*) In the third method, detection is indirect through the use of the labeled antibody. (*d*) This illustrates a homogeneous competitive method in which the properties of the label itself (i.e., the signal) changes. (Adapted from G. Kersten and J. Westdijk, Immunoassays, in W. Jiskoot and D. J. A. Crommelin (Eds.), *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005.)

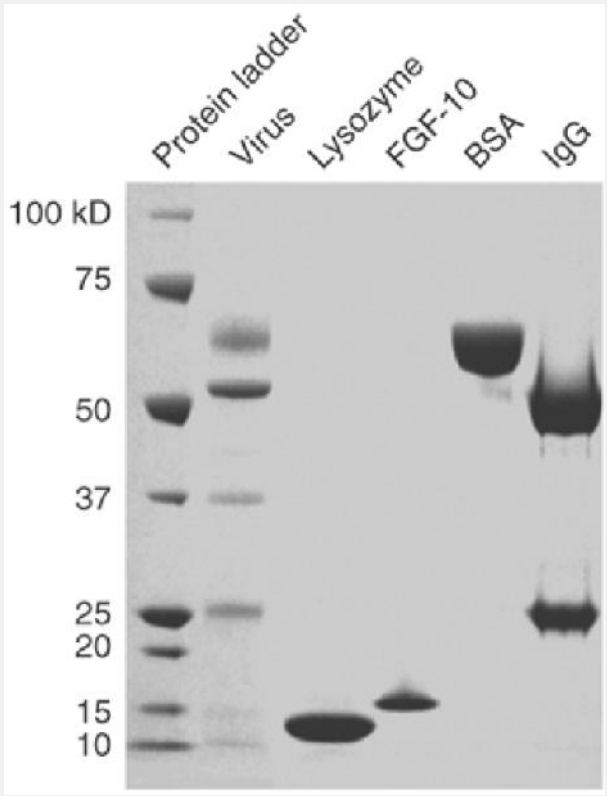

Fig. 21-7. SDS-PAGE of representative biopharmaceuticals under reducing conditions employing a gradient gel. As expected, the virus produces multiple protein components and the IgG, two bands corresponding to the heavy and light chains. Molecular weights can be estimated by comparison to the known standards present in the left-most lane.

Because nucleic acids carry a strong negative charge due to their phosphate groups, they can also be electrophoresed. In this case, however, a gel based on agarose is usually employed. The migration of DNA and RNA is dependent on their structure. Small pieces of double-stranded DNA or RNA generally migrate as a function of their radius of gyration, but larger species such as supercoiled DNA or single-stranded

nucleic acids which can internally base pair and therefore form tertiary structure can behave in quite complex ways (Fig. 21-8). Nucleic acid agarose gels are usually stained with a dye such as ethidium bromide which intercalates between the bases, causing it to become highly fluorescent. The high separatory capacity of agarose gel electrophoresis (polyacrylamide can also be used) makes this an especially important analytical tool.

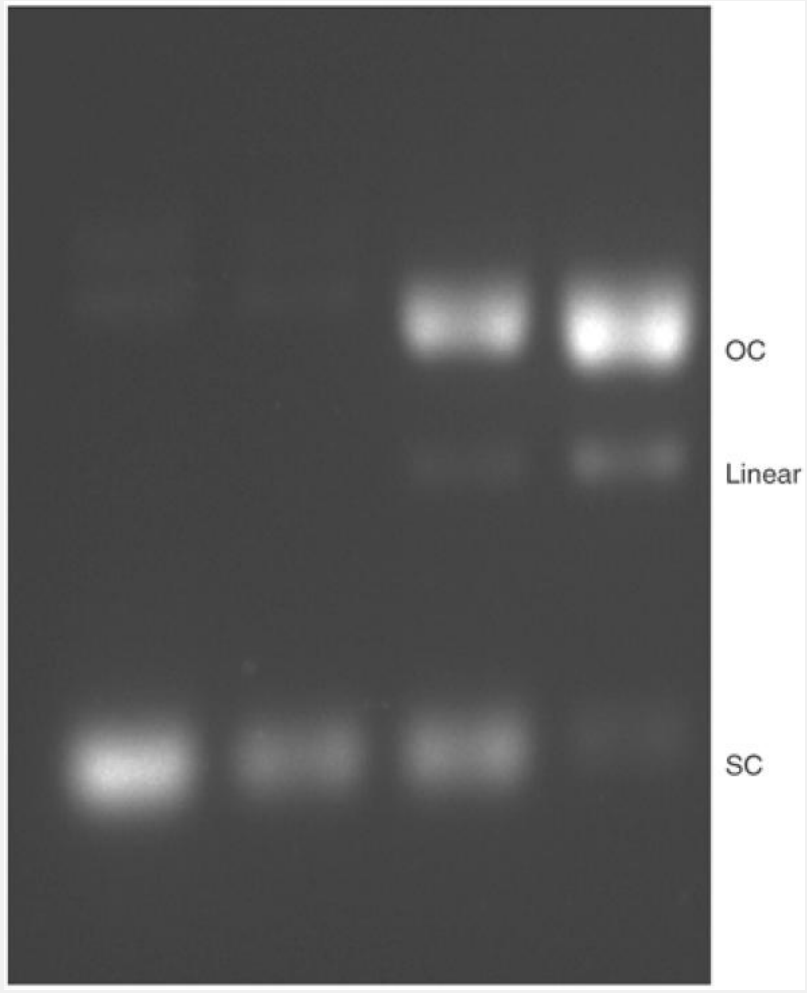

Fig. 21-8. An agarose gel of a supercoiled plasmid (sc) (left) which as been subjected to long-term storage with consequent development of open circle (oc) and linear forms (right).

A third variant of electrophoresis of great importance is isoelectric focusing. Here, compounds known an ampholytes (mixtures of charged peptides) are included during electrophoresis. The externally applied voltage causes these charged agents to form a pH gradient within the gel. This causes the target protein to migrate to the point in the ampholyte gradient when its charge is neutralized, that is, its isoelectric point. Thus, this method allows one to measure the pH at which a protein is charge neutral. Proteins differing by a single charge (such as seen during a deamidation event, see below) can be distinguished. High concentrations of urea are often included during gel isoelectric focusing to reduce extraneous protein/protein interactions.

Electrophoresis can also be conducted in small capillaries containing an electrolyte (capillary zone electrophoresis or CZE). Like gel electrophoresis, species are generally separated on the basis of their size to charge ratio. In this type of analysis, the motion of the buffer solution itself under the influence of the electromagnetic field (electroosmotic flow) often exceeds that of the electrophoretic migration of the sample and this contributes significantly to the separation. A variety of detection methods are available and are usually instituted through a clear area of the capillary. The most commonly used methods are UV/VIS absorption and fluorescence (either through the intrinsic fluorescence of the macromolecular such as protein tryptophan fluorescence (see below) or an extrinsic fluorescent group introduced through chemical modification). The output of the capillary can also be fed to a mass spectrometer for

molecular weight analysis. Isoelectric focusing can also be performed in a capillary format. Although more commonly used for proteins, capillary electrophoresis can also be used for nucleic acids. In a combination of the gel and capillary approaches, a separatory gel matrix can be included within the capillary. Electrophoresis has also been used to measure the zeta potential of macromolecules. This is discussed in Chapters 15 and 16.

*High-Performance Liquid Chromatography*46

In addition to electrophoresis, HPLC in its myriad forms is the most common analytical technique used to characterize macromolecules. The various types of separations available for HPLC are similar to those described above for the different types of preparative chromatography. The major differences are the use of small diameter columns packed with smaller particles that permit the use of high pressure (more than 1000 atmospheres) and consequent rapid flow rates and separation times (typically minutes). A wide array of detection methods is available including UV/VIS absorption, fluorescence, vibrational spectroscopies, electrochemical methods, and mass spectrometry.

Probably the most frequently used form of HPLC is the reversed-phase mode (RP-HPLC). In this method, the packing material is derivatized with alkyl or other apolar substituents. The molecules to be separated interact with the apolar moieties through hydrophobic sites on their surfaces. They are then eluted with an apolar solvent like acetylnitrile or propanol, often in a step or gradient mode. Under optimized conditions, the time of elution (the “retention time”) may be taken as highly characteristic of the analyte and can be used to establish its identity and structural integrity. A common use of RP-HPLC with proteins is the construction of “peptide” maps (Fig. 21-9). In this important procedure, a protein is digested into peptide fragments of defined size by one or more proteases with their molecular weight determined by online mass spectrometry. This permits the chemical (primary) structure of a protein to be rigorously established and is routinely used to confirm a protein's identity and the presence of chemically altered residues. In general, peptides tend to elute on the basis of their relative polarity. The elution behavior of proteins themselves is much more complex due to their secondary, tertiary, and quaternary structure, but it is assumed that the presence of apolar binding sites plays a key role in their elution behavior as well as conformational changes produced by the eluting solvent.47

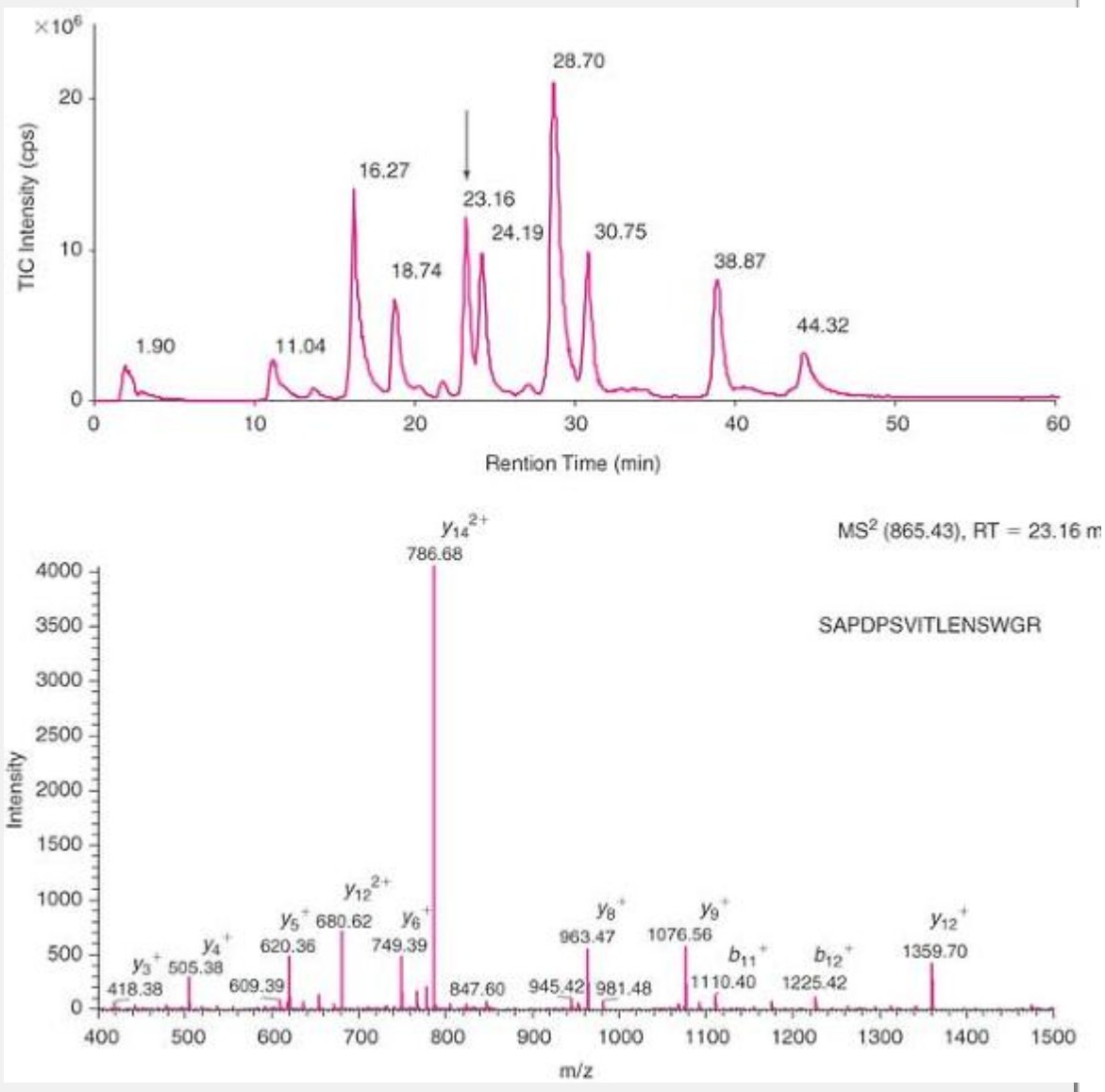

Fig. 21-9. Peptide map of a mutated Ricin-A Chain after digestion by trypsin and analyzed with LC-MS/MS. The peptides were separated on a reverse-phase column C_{18} with gradient elution from 0% to 80% acetonitrile in water (containing 0.06% formic acid) at 10 µL/min, and the separated peptides were detected by mass spectrometry (upper figure). The lower figure shows a product ion spectrum which can be used to identify the sequence of the peptide shown in the upper right corner. The fragments with chargers on the N-Terminal side are called "b" ions and the fragments with chargers on the C-terminal end are called "y" ions.

There are many additional forms of HPLC that are used as analytical tools in macromolecular analysis. Especially important are size exclusion (SEC), ion exchange and bioaffinity chromatography. SEC is of exceptional importance since it is used to detect changes in size. Thus, degradation that results in significant decreases in size or increases in mass such as that produced by limited association or more extensive aggregation is often first detected by this sensitive technique. One problem with its use is that dilution occurs during separation and thus the concentration at which separation occurs is not that of the initial sample. Furthermore, interactions of proteins with a column's matrix can significantly alter elution behavior. This makes estimates of size sometimes difficult. Inclusion of agents such as urea and high

salt concentration can be used to minimize this problem. Nevertheless, this method provides the most frequently used criteria for detection of the crucial phenomenon of protein aggregation. The method can also be used for nucleic acids and even viral particles if highly porous resins are used. With the advent of high pressures and appropriate packing materials, it is now possible to obtain analytical separations in a few minutes (and even seconds in certain circumstances) making HPLC-based methods of both widespread use and significant analytical importance.
P.529

Ultraviolet/Visible Spectroscopy

A variety of spectroscopic methods are widely used to characterize biomolecules. An introduction to the general topic of spectroscopy is presented in Chapter 4 and it is highly recommended that this section be reviewed before perusing this section. This chapter is restricted to aspects of these techniques that are particularly relevant to biomolecules in a pharmaceutical context.

Absorption spectroscopy in the ultraviolet and visible regions is a very versatile technique widely used with both proteins and nucleic acids.[48,49] The technique has shown a recent resurgence due to the availability of diode array detection. Conventionally, absorption spectra were produced by scanning a moveable monochromator (a light dispensing element) through a sample contained in a cuvette (a visible and/or UV transparent rectangular sample holder of path length 0.01–10 cm, most commonly 0.1–1 cm) with detection by a photomultiplier tube. In contrast, in a diode array instrument, all wavelengths of light are put through the sample simultaneously and a spectrum created after absorption of the light. The resultant spectrum is then projected onto a diode array for detection purposes. Most importantly, by using mathematical fitting techniques to interpolate between the individual wavelengths detected by the diodes (typical spaced at 0.5–2 nm), highly resolved spectra can be obtained after derivative analysis to produce a resolution on the order of ±0.01 nm.

In proteins, there are two major intrinsic chromophoric groups to consider. In the far UV region (175–220 nm), there are three electronic transitions observed because of peptide bonds with a broad peak seen at 185 to 195 nm. Although this region contains information about a protein's secondary structure, it has rarely been used for this purpose due to optical interference by most substances. The high wavelength tail is often used, however, to detect proteins during various forms of chromatography.

In contrast, the near UV region is used for a wide variety of purposes. This portion of a protein absorption spectrum (240–310 nm) is dominated by the $\pi \rightarrow \pi^*$ transitions of the three aromatic amino acid side chains. Phenylalanine (Phe) manifests a weak peak with marked vibrational structure between 250 and 270 nm, tyrosine (Tyr) a stronger, pH-dependent multicomponent peak from 250 to 290 nm, and tryptophan, the strongest absorbing side chain (another multicomponent peak) from 250 to 300 nm. In the case of proteins, the broad overlapping nature of these three contributions results in a broad peak centered between 277 and 287 nm (primarily from Trp and Tyr) with weak undulating bumps in the spectrum below 270 nm (due to the vibrational fine structure of the Phe contribution) and a marked shoulder for Trp at approximately 290 nm. Weak contributions from His and disulfide bonds can also occasionally be seen. The second or fourth derivative of a protein's spectrum dramatically brings out the underlying contributions usually in the form of 6 peaks (3 Phe, 1 Tyr, 1 Tyr/Trp, and 1 Trp) (Fig. 21-10). As indicated above, these peaks can often be localized to within 0.01 nm, providing a highly distinctive spectrum for each protein. Some additional information is possible since the Phe residues are usually buried in a protein's interior, most Tyr are interfacial, and the Trp residues dispersed throughout the protein's matrix. The important point here is that the position (wavelength) of each of these peaks is sensitive to the polarity of their environment. The general rule is that as the
P.530

environment of an aromatic amino acid's side chain becomes less polar (more hydrophobic), its wavelength is shifted to a higher wavelength (note that this effect is opposite to that seen in a protein intrinsic fluorescence experiment as will be discussed below). Thus, as the structure of a protein is

altered, to the extent that one or more aromatic side chains experiences a subsequent change in its immediate environment, changes in the position of the derivative absorption peaks provide a measure of conformational change (and potentially physical degradation). The folding and unfolding of a protein as induced by temperature, pH, or a potential unfolding agent like urea, guanidinium hydrochloride, alcohols, chaotropic salts, and detergents can be simply followed by this method if their optical properties do not interfere with the measurements. Data obtained by absorption measurements as a function of multiple variables such as temperature and pH can be summarized and visualized by a method known as the “empirical phase diagram (EPD).” In this approach, the six peak positions are used as components of a vector at any particular set of variables (e.g., T, pH). If colors are then assigned to the major components, a map of temperature versus pH displays different colored regions corresponding to different physical states of the protein (Fig. 21-11). For more information about this method, see reference (50). By far the most common use of protein UV spectroscopy, however, is to determine their concentrations using Beer's law (Chapter 4), which states that the concentration is linearly proportional to the absorbance with a constant of proportionality known as the extinction coefficient (ε, $A = \varepsilon cl$). Extinction coefficients, which are characteristic of each molecule, can be determined by dry weight or amino acid analysis or calculated from the aromatic amino acid content from a number of empirical algorithms.[51,52]

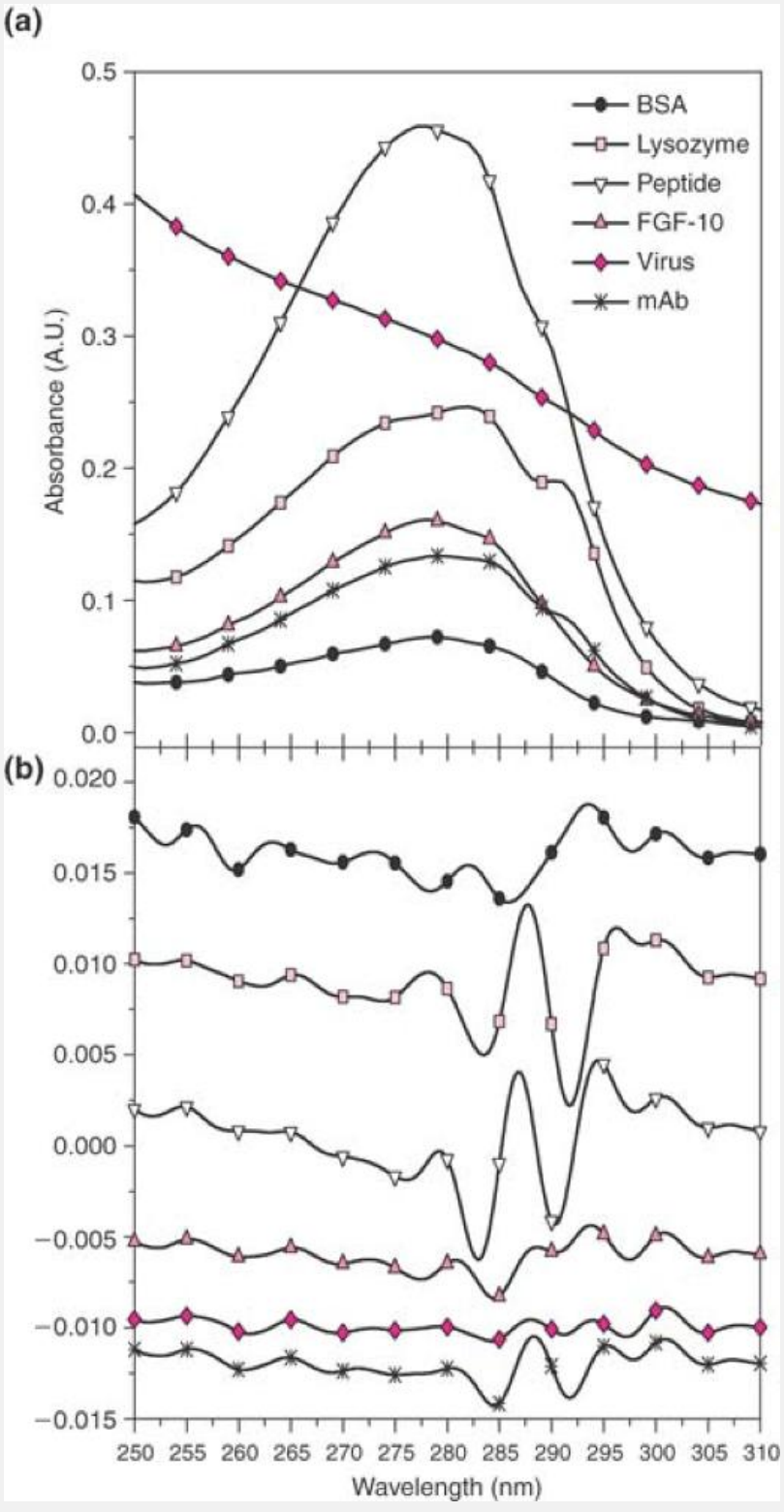

Fig. 21-10. Zero-order absorption spectra of four proteins, a peptide and a viral particle (upper panel). Note that all display rather broad featureless spectra. When their second derivative is calculated, however, usually six peaks are seen corresponding to the three classes of aromatic side chains (bottom panel). The second

derivative spectra in the lower panel are displaced for clarity. See text for further discussion of the interpretation of such spectra.

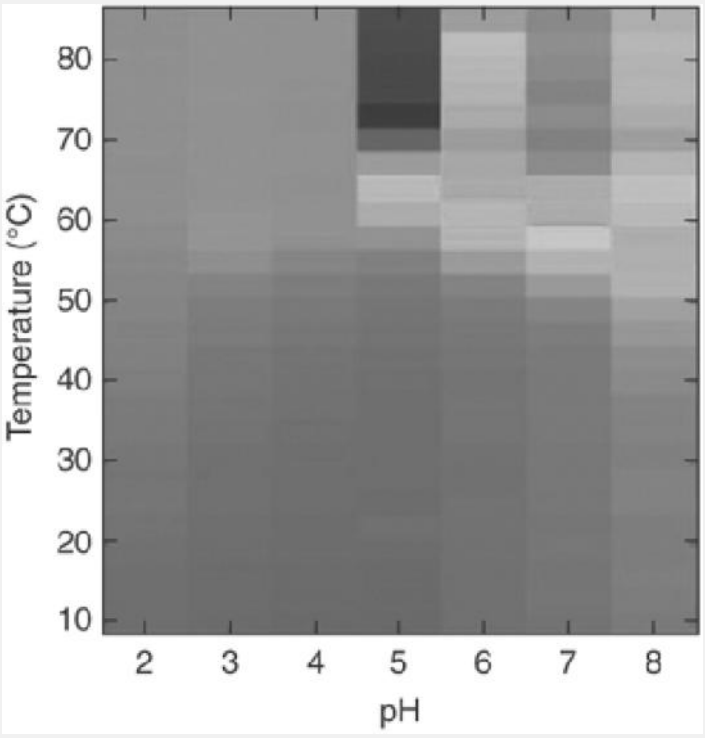

Fig. 21-11. An empirical phase diagram (EPD) based on the peak positions of second derivative UV absorption spectra. At each T/pH condition, the protein is represented as a six-dimensional vector which is truncated to the three largest contributions. The regions with different shaded characteristics represent different structural states of the protein. Such diagrams are usually shown in color as described in the text.

A number of extrinsic chromophores may also be present in proteins. These include metals such as copper or iron as well as lanthanides (with Tb and Co the most commonly employed), which can be used as calcium analogues. Prosthetic groups such as flavin-adenine dinucleotide (FAD), flavin mononucleotide (FMN), NAD, rhodopsin, pyridoxal phosphate, and heme groups among others, all provide strong spectra in the UV/VIS region and are sensitive to their local environment. They can therefore be used as biological sensors of a variety of phenomena such as redox state, O_2/CO_2 binding, and light effects.

*Fluorescence*53’54

Probably the most versatile spectral technique for macromolecular structural analysis is fluorescence spectroscopy. When a chromophore such as the indole side chain of tryptophan in proteins is raised to an excited (singlet) state, rather than return to the ground state through internal conversion processes as in an absorbance measurement, it can do this by the emission of a photon. This constitutes the phenomenon of fluorescence. If this emission is from a triplet state, the process is known as phosphorescence (the combination of the two is called luminescence). For brevity, we will not be concerned with the latter here although the technique of phosphorescence can be quite useful in the analysis of proteins. In the case of fluorescence, the (relatively) long periods of time spent in the excited states (10^{-3}-10^{-9}s versus <10^{-15}s for absorbance) allow various types of interactions with this state. This makes fluorescence usually quite sensitive to the fluorophores' immediate environment. The spectrum

of this emission is always at longer wavelengths than the absorption band(s) because prior to emission of fluorescence photons, energy is lost as the excited state returns to its lowest vibrational energy level. As a first approximation, the absorption and emission spectra are mirror images of one another. In addition to emission spectra, the lifetime of the excited state (*r*; the time it takes fluorescence to fall to 1/e of its initial value) can also be measured. The amount of emission can be measured either in terms of its quantum yield (the number of photons emitted divided by the number absorbed) or by simple intensity changes at a fixed wavelength. The latter is approximately proportional to the quantum yield. Fluorescence is usually measured at right angles to the exciting light, with the emission monochromoter scanned. Alternatively, the excitation monochromoter can be varied and emission monitored at a fixed wavelength to produce a version of the absorption spectrum known as an excitation spectrum. Lifetimes are measured by either exciting with single pulses of light and measuring their emission decay or determining shifts in the phase of the emitted photons after modulation of the exciting light. Both techniques can measure more than one lifetime component (up to three or four) by deconvolution methods or display distributions of

lifetimes. Most lifetimes of natural fluorphores are on the order of a few nanoseconds (Trp [~2.6 nsec], Tyr [~3.6], Phe [~6.4], NADH [~0.4], etc.). In general, low quantum yields correlate with short lifetimes.

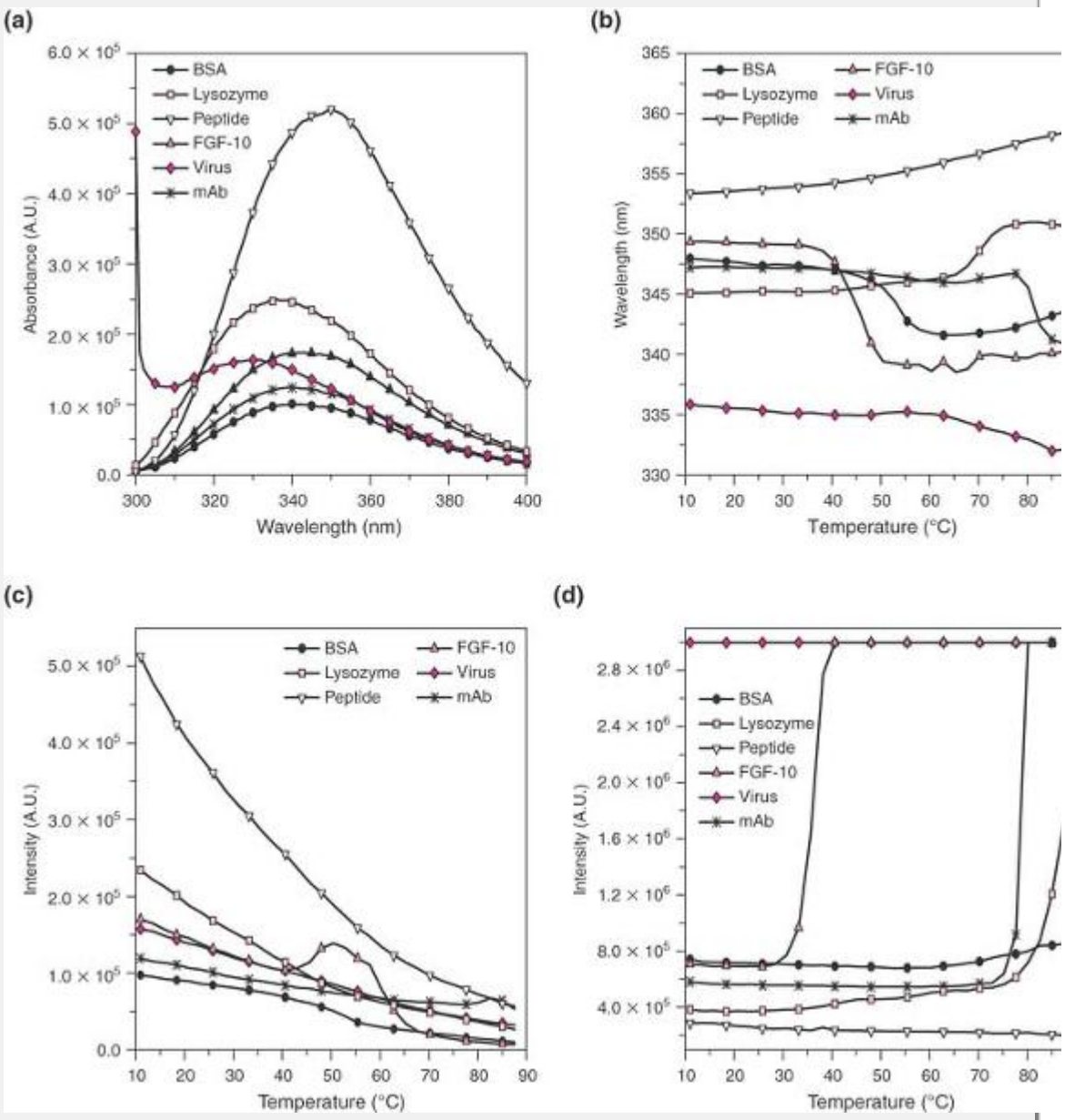

Fig. 21-12. Intrinsic fluorescence spectra of biopharmaceuticals. (*a*) The position of the emission maximum (upon excitation at 295 nm) varies from approximately 330 to 350 nm in these examples. Emission at 350 nm indicates that the tryptophan residues are on average highly exposed to the solvent. As the emission maximum decreases, increased burial of the indole side chains in indicated. (*b*) The effect of temperature on the position of the emission maximum and (*c*) its intensity. Note the appearances of transitions in all the macromolecules with the exception of the peptide. (*d*) Light scattering at 295 nm as a function of temperature acquired simultaneously with the intrinsic fluorescence data. The high value seen with the viral particle directly reflects its large size compared to the individual proteins and the peptide.

The intrinsic fluorescence spectra of proteins tend to be dominated by tryptophan emission unless this residue is absent. Under such conditions, the emission of tyrosine and or phenylalanine can often be seen. Of special importance is the observation that this fluorescence is very sensitive to the environment

of the endogenous indoles (Fig. 21-12). Thus, various kinds of conformational changes that range from the very subtle to complete unfolding can usually be followed by this method. The position of protein Trp containing emission spectra varies from 310 to 320 nm (completely buried indole

side chains) to more than 350 nm (totally exposed to the aqueous solvent). Other phenomena such as ligand binding and subunit association and dissociation can also often be detected by changes in intrinsic fluorescence. In a particularly elegant use of this approach, single Trp residues can be placed at many positions throughout the structure of a protein by site-directed mutagenesis (it may be necessary to remove some tryptophans if the target protein contains more than one) and by measuring peak positions to estimate Trp exposure. This information can then be used to generate actual 3-D structures using additional modeling considerations.

Studies of the fluorescence of proteins are not limited to the aromatic amino acids. It is also possible to add an extrinsic probe, which either covalently or noncovalently binds to a particular site on a protein. Common covalent fluorescent labels include molecules like dansyl chloride or fluorescein isothiocyanate. Covalent labels are available which bind to a variety of protein reactive sites such as amino, carboxyl, and sulfhydryl groups. Such probes are typically highly fluorescent and their spectral properties highly sensitive to their environment. Examples of fluorescent probes which are used to bind noncovalently to proteins include 8-anilino naphthalene sulfonic acid (ANS) and its dimeric analogue bis-ANS. Probes such as this are usually assumed to bind to apolar sites on proteins although the negative charge on ANS may also result in electrostatic interactions. One use of ANS is to detect molten-globule (MG) states in proteins. In such states, as the tertiary structure begins to be disrupted, compounds like ANS can interact with the protein and their normally solvent quenched emission can be relieved with enhanced fluorescence expressed as well as blue shifts in wavelength emission maxima. Another use of noncovalent probes involves the detection of protein aggregation. When proteins self-associate and then aggregate, they often form intermolecular β-structure. Certain dyes (e.g., Congo Red, Thioflavin T) can bind to such structures with a change in their fluorescence or absorption spectra.

Nucleic acids lack significance fluorescence (with the exception of a modified base in tRNA). Noncovalent probes, however, have been extensively used in their characterization and analysis. Some planar dyes can intercalate between nucleic acid bases, whereas others can bind within the grooves of the helix. Upon such interactions, their fluorescence can dramatically increase. This phenomenon has been used to measure DNA and RNA concentration, analyze the binding of other substances to nucleic acids through dye displacement (Fig. 21-13), and follow the behavior of DNA microscopically among other applications. Fluorescent dyes are also commonly used to study lipid bilayers and cell membranes. Using probes that enter into the bilayer, the fluidity of membrane interiors and associated phase changes can be monitored by changes in the spectral properties of certain dyes, often using temperature to induce phase transitions.

There are many additional applications of fluorescence to macromolecular systems. Only three are briefly considered here. Quenching of protein Trp fluorescence by extrinsically added solutes is a commonly used approach to analyze the accessibility of these residues to the aqueous solvent. Certain low molecular solutes can quench indole fluorescence either by a dynamic collisional process (dynamic quenching) or by forming a complex with the side chain, preventing it from reaching an excited state (static quenching). Some commonly used quenchers are 0_2, acrylamide, I^-, and Cs^+. The different charge characteristics of the quenchers permit them to probe different electrostatic environments. The dynamic quenching process can be described by a relationship known as the Stern-Volmer equation:

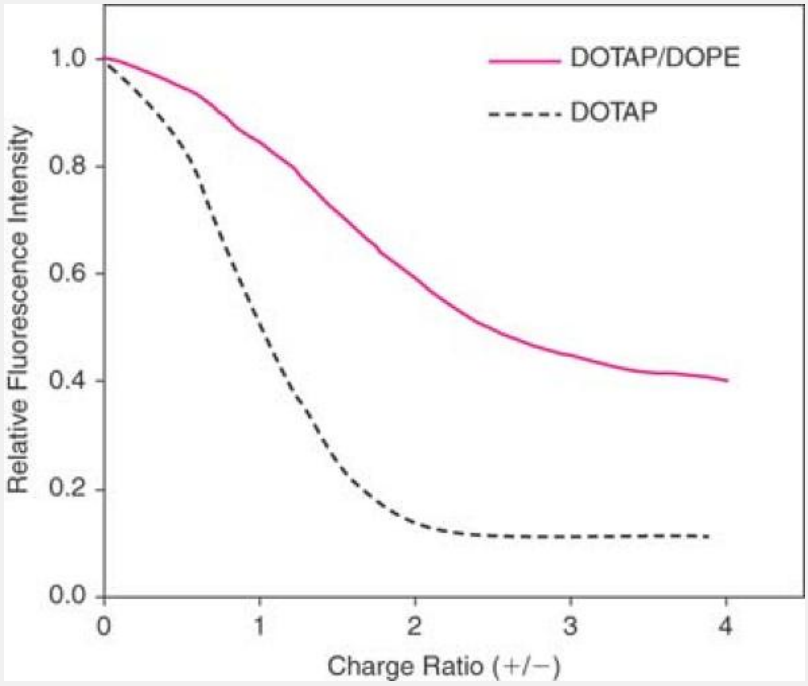

Fig. 21-13. An example of the use of a fluorescent dye to measure complex formation between DNA and cationic lipids. The dye ethidium bromide is intercalated between the bases of a DNA plasmid where it becomes highly fluorescent. When cationic lipids are added, the dye is displaced and its fluorescence reduced. If the charge density of the cationic lipid is reduced by the addition of a neutral helper lipid (DOPE, dioeyl phosphatidylcholine), its efficiency in displacing the dye is reduced. Formulations containing such helper lipids are actually more efficient at transfecting target cells than those with cation lipids alone.

$$\frac{F_o}{F} = 1 + K[Q]$$

or

$$\frac{F_o}{F} - 1 = k_q \tau [Q]$$

where F_o is the fluorescence (or quantum yield) in the absence of quencher, F the fluorescence in the presence of quencher, Q is the concentration of quencher, and K is a constant known as the Stern-Volmer constant. If the quenching process is entirely dynamic, a plot of F_o/F versus [Q] will yield a straight line of slope K or $k_q\tau$. Curvature in such plots can be introduced by hetereogeneity in the environment of multiple Trp residues or the presence

of static quenching. If the lifetime (τ) of the indole is known, the rate constant (k_q) for the collisional process can be evaluated. This can be used to characterize changes in structure or interestingly, the dynamic behavior of proteins in terms of the ability of the protein matrix to permit the transport of the quencher to the indole side chains. We will consider below other ways, including some that involve fluorescence, of studying the intramolecular dynamics of proteins. If no Trp is present in a protein, this method may also be used to study the quenching of Tyr residues. An ultraviolet absorbance technique can be used to acquire related information. In this case, simple cations of various sizes are added in increasing amounts to proteins and the positions of the derivative peaks from Trp, Tyr, and Phe are monitored. In this case, the shifts are due to the formation of cation-π interactions of cations such as K^+,

Cs^+, and Li^+ with the negatively changed π electrons of the aromatic rings. Since smaller cations diffuse more easily into the protein interiors, this method can also be used to analyze protein dynamics.55

Another method in wide use is known as singlet–singlet energy transfer (or fluorescence resonance energy transfer or FRET). If the emission spectrum of a fluorophore (the "donor") overlaps the absorption spectrum of a second fluorophore (the "acceptor"), then when the donor is excited and the donor and acceptor are close to one another, under certain circumstances the acceptor will emit radiation as the donor is quenched. When this is *not* due to the trivial reemission of an absorbed photon, the process occurs as a result of a resonant interaction between the emission process of the donor and the absorption process of the acceptor. The efficiency of this transfer process is a function of the spectral overlap, the relative orientation of the two fluorophores and the distance between them. This distance (R) is given by:

$$R = R_0 \left(\frac{1 - E}{E} \right)^{1/6}$$

where E is the efficiency of the transfer process and R_0 is the characteristic transfer distance that corresponds to R where E = 50%. R_0 is a function of the degree of spectral overlap, the refractive index between the donor and acceptor, an orientation factor (κ^2), which depends on the relative orientation of D and A, and the quantum yield of the donor. Methods exist to estimate the orientation factor but a value of 2/3 for a random orientation is usually used.

This method is most commonly used by placing either covalently or noncovalently specific fluorophores with the proper spectral properties (especially spectral overlap) at single locations either within a macromolecule or at sites in different molecules that are sufficiently close (<80 Å) that efficient resonance energy transfer can occur. Thus, relatively accurate distance estimates can be determined by this method. Numerous systems have been examined by variations of this technique and its utility is well-established for mapping a wide variety of structural features of single molecules, molecular complexes such as viruses and ribosomes, and the surface of cell membranes. FRET has also frequently been used in both static and kinetic modes to study nucleic acids and their complexes.

A third important application of fluorescence involves the use of polarized radiation. This is based on the principle that there is preferential absorption of light when a chromophore has its transition dipole(s) parallel to that of the exciting light. Thus, if polarized light is used to excite a randomly oriented collection of fluorophores, those transiently oriented parallel to the exciting light will preferentially absorb light. If the "photoselected" molecules rotate within their excited state lifetime, the emitted light will be depolarized to some extent. Thus, the motion of fluorophores and the molecules to which they are attached can be analyzed by this method. Depolorization is typically measured in terms of a quantity known as the fluorescence anisotropy (r), which is defined as:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

where $I_{||}$ and $I_{\perp}$ are the intensities of the emitted light oriented parallel and perpendicular to the exciting light. Experiments can usually be performed in one of two ways. In steady state studies, the anisotropy or the depolarization ratio ($I_{\perp}$ - $I_{||}/I_{\perp}$ + $I_{||}$) is measured. By varying the temperature or viscosity, it is possible to calculate the rotational correlation time (ρ) of the macromolecule to which the fluorophore is attached. Alternatively, the target fluorophores can be excited with single photons of polarized light and the anisotropy of the emitted light detected. Because the anisotropy (A(t)) decays exponentially with time, molecules that emit later have more time to rotate:

$$A(t) = A_0 e^{-3t/\rho}$$

where the rotation correlation time ρ is just the time it takes a molecule to rotate 1/e of a complete rotation. Multiple rotation modes can often be resolved by deconvoluting the experimental data into individual exponential components. Thus, it is frequently possible to resolve internal dynamic motions of macromolecules from their overall motion. Polarization measurements have also been widely used to study ligand binding and protein/protein interactions. If a fluorescent label is attached to the smaller component, its polarization increases as it becomes part of the larger complex. There are many variations on such experiments, but the results can be quite sensitive to the presence of the interactions.

We have only touched on the use of fluorescence techniques for biological molecules here. A wide variety of experimental methods are based on this principle. These include fluorescence microscopy, single molecule fluorescence, fluorescence photobleaching and recovery, and fluorescence correlation spectroscopy. The interested student is referred to the comprehensive text by Lakowicz for further information.54

Circular Dichroism

The physical basis of circular dichroism (CD; see references 56 and57) is different from that of simple absorption and fluorescence and as such can provide somewhat different information: light can be thought of as composed of two opposite circularly polarized components. If one of these components is greater or less than the other due to differential absorption and they are combined, light that is elliptically polarized is

produced. It is the angle of rotation of the long axis of such an ellipse that is measured in a CD experiment (i.e., the “ellipticity”). CD is seen only in absorption bands, thus requiring appropriate chromophores. CD signals are produced by the interaction (technically, through the “dot product”) of electronic and magnetic absorption processes. This is, in fact, the general origin of optical activity. Electronic absorption can be thought of as a unidirectional displacement of charge, whereas magnetic absorption can be represented as a light-induced current loop. When a vertical motion of charge acts on such a circular displacement, a helical charge distribution is produced. As we will see in a moment, this is especially important for CD analysis of proteins and nucleic acids.

According to the above, CD can only be produced when the local environment of a chromophore is asymmetric. In biomolecular systems, there are at least three situations in which this is seen. In the first, transitions could involve electrons near the α-carbon atoms of amino acids in proteins. Because there are no major absorptive chromophores here, however, any such signals are quite weak. Second, the tertiary structure of a macromolecule could place relatively symmetric absorptive molecular groups into asymmetric environments. Third, helices can facilitate a helical flow of charge thus producing relatively large optical activity if there is an appropriate absorptive chromophore. It is this last situation that has received the most attention and we will begin our discussion of the CD of proteins and DNA here.

Proteins have several chromophores of potential interest from a CD perspective. These are the peptide bond, certain side chains (especially the aromatic side chains and disulfide bonds), endogenous chromophores such as heme groups, and extrinsically added chromophores whose optical activity changes or is induced when they are added to proteins. The CD of the peptide bond consists of a band near 222 nm ($n \rightarrow \pi^*$) and a signal, which is split into two parts through interactions between transitions at 200 to 210 nm ($\pi \rightarrow \pi^*$, $\parallel$) and 191 to 193 nm ($\pi \rightarrow \pi^*$, $\perp$) (Fig. 21-14). Right-handed alpha helices produce a distinct CD spectrum with negative peaks at approximately 208 and 222 nm and a positive peak near 192 nm. β-sheets manifest a weaker signal (β structure can be thought of as distorted helices) at 215 to 218 nm and a positive peak at about 195 nm. Beta turns give a number of weak signals in the same region depending on their type. A left-handed α-helix results in a spectrum that is approximately a mirror image of the right-handed version. Cross-β-structure and beta-trefoils produce negative signals at 210 to 215 and 203 to 208 nm, respectively. Lastly, disordered structure typically gives a peak with a 195 to 200 nm minimum and often a weak positive signal near 230 nm. Because of the distinct nature of the CD spectra of these different types of secondary structure, a protein's CD spectrum can be used to estimate fairly accurately its secondary structure content. The basic idea is a simple one. One can extract inherent values for the ellipticity of the various types of secondary structure from a library of proteins of known secondary structure content (determined by x-ray crystallography or NMR). These values will reflect typical effects of 3-D structure, the contribution of side chains, and the length and distortion of regions of secondary structure as well as other factors. This data serves as a basis from which to fit unknown spectra and provide secondary structure content estimates. Such an analysis does not provide absolute values, but the fractional content determined and is often good to 2% to 3%. The method is especially powerful when used to monitor changes in secondary structure with

very subtle changes in helix and β-sheet content detectable. Note the specificity of analysis of secondary structure when CD spectra

are obtained in the 180 and 250 nm range. In contrast, UV absorption and fluorescence in the near UV region are primarily sensitive to changes in tertiary structure although they will reflect indirectly secondary structure alterations. CD does, however, offer tertiary structure information when used between 250 and 300 nm where the aromatic side chains and disulfide bonds become the absorbing chromophores. The protein spectrum in this region is quite complex, consisting of a series of positive and negative peaks. These are primarily produced by the induction of optical activity in these side chains although they do possess some weak, intrinsic optical asymmetry. Attempts to derive specific structure information from the details of such spectra have largely been unsuccessful. Changes in the CD spectra of proteins in the 250 to 300 nm region, however, are quite useful in the same way that UV absorbance and intrinsic fluorescence spectra are employed to detect changes in tertiary structure. These signals are, however, significantly weaker than the far UV peaks and either higher concentrations or longer cuvette path lengths must be used. In a number of cases, other intrinsic chromophores such as heme or rhodopsin groups or extrinsically added dyes can produce strong CD signals associated with their absorption bands. If detailed structural information is available, it may be possible to further use such spectra to provide additional structural information. There also exist a number of hybrid versions of CD such as vibrational, magnetic, and fluorescence-detected circular dichroism, which are useful in specific indications. In general, however, the most common use of protein CD involves the monitoring of protein conformational changes (in terms of changes in either secondary or tertiary structure, or both). Alternatively, the effect of ligand binding either through the induction of conformational changes in proteins or induced optical activity in the ligand is another common use. Using the latter method, binding constants and stoichiometries can often be determined. If temperature is varied, the enthalpy, entropy, and heat capacity of the binding process may also be determined using a Van't Hoff analysis.

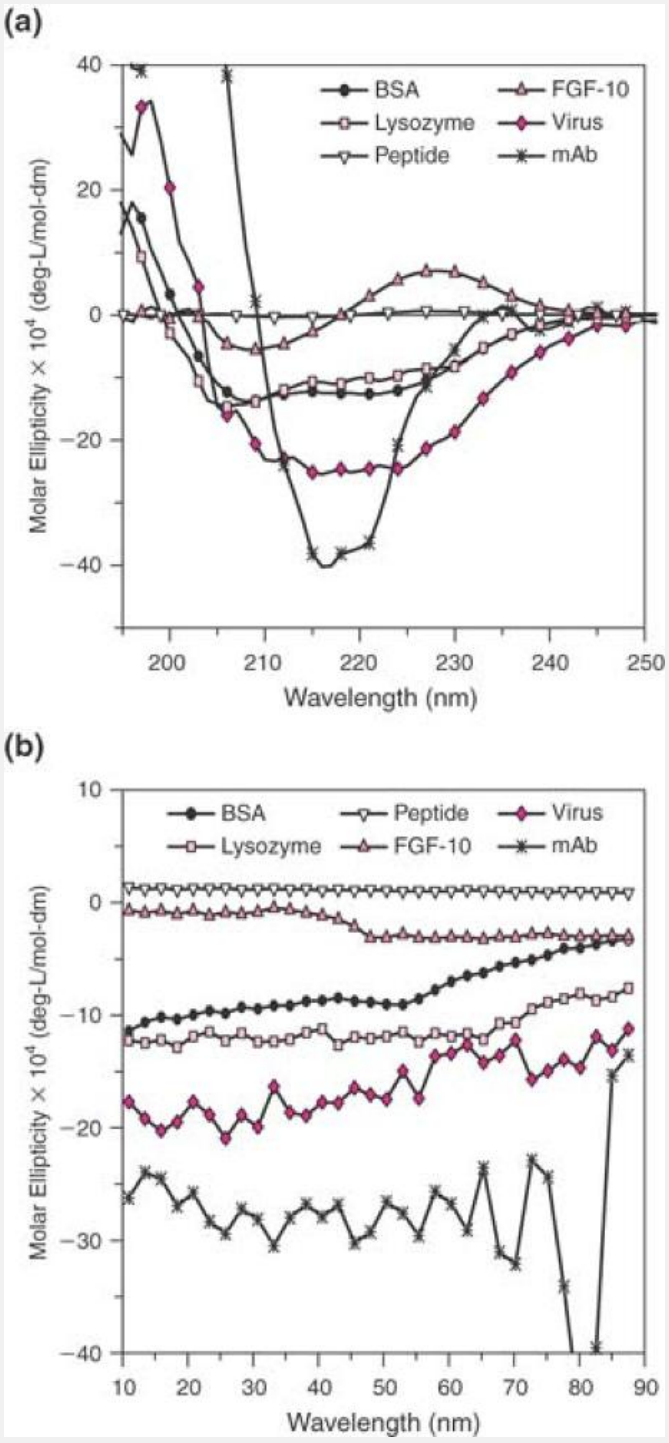

Fig. 21-14. Circular dichroism (CD) spectra of representative biopharmaceuticals (*a*) with (*b*) accompanying thermal melting curves. A strong double minimum is seen at 208 nm and 222 nm for the α-helix rich BSA and lysozyme. The weak β-structure minimum at 217 to 218 nm for IgG is more difficult to see on the same scale. Note the positive peak for the peptide between 225 and 230 nm which is often assigned to loosely coiled peptides.

Because of their helical nature, nucleic acids also produce strong CD signals. In this case, the CD arises from the nucleotide bases and their absorption between 200 and 300 nm. The CD spectra of the different forms of the nucleic acids are quite distinct from one another due to differences in interactions between the bases (Fig. 21-15). For example, the A form of DNA (11 base pairs bp/turn) produces a spectrum with a maximum near 270 nm and a minimum near 210 nm. The spectrum of A-RNA has a spectrum similar to A-DNA but is shifted to 10 nm lower wavelengths. B-DNA (10 bp/turn) has a less intense spectrum that is similar in shape but the peaks are slightly shifted. Z-DNA (12 bp/turn, but left-handed in contrast to the A and B forms) is roughly a mirror image of the A and B forms. Depending upon the actual sequence of the nucleic acids, the spectra are subtly different but most importantly highly sensitive to changes in structure. Thus, the melting of ds DNA or RNA can easily be followed by this method. If small pieces of DNA (oligonucleotides) which contain specific binding sites are examined, the binding of proteins to DNA can be analyzed by CD changes. When dyes, drugs, and delivery agents bind to DNA, they often display quite marked induced CD or changes in the spectrum of the nucleic acid (Fig. 21-15). This can be used to study their interaction. If DNA becomes highly compacted which can occur when its charges are neutralized, quite unusual spectra can be produced which are diagnostic of unique forms of condensed DNA.58

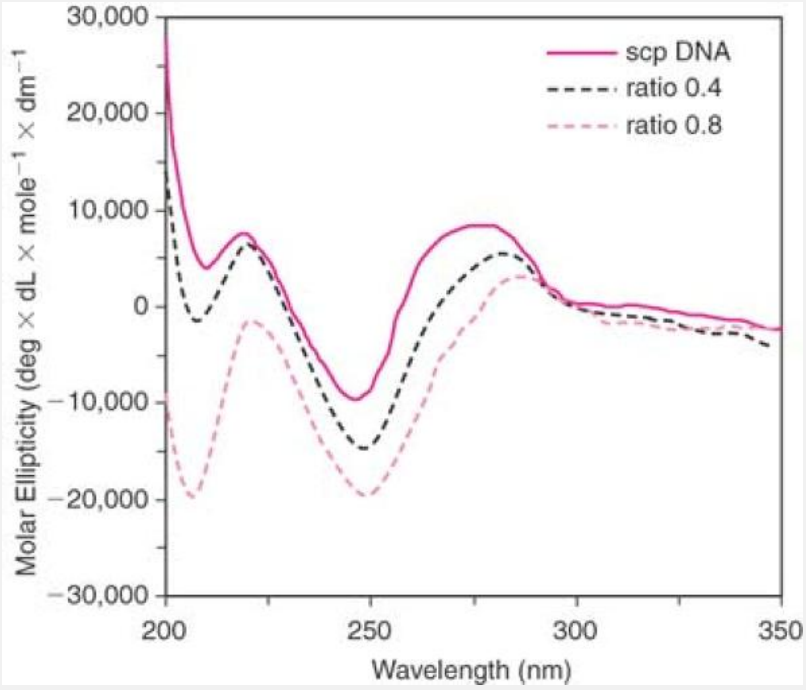

Fig. 21-15. CD spectra of pDNA and the DNA in the presence of a cationic polymer complex at different charge ratios (+/-). The CD spectrum of the DNA alone shows that it is in the B-form. The presence of the cationic polymers causes the structure of the DNA to change. A change in the CD spectrum of the DNA is seen because the interaction between the nucleic acids bases is altered, thus perturbing the helical nature of the DNA.

Vibrational Spectroscopy

The secondary and tertiary structure of both proteins and nucleic acids can also be analyzed by vibrational spectroscopy. Both infrared[59,60,61,62,63] and Raman[64,65] spectroscopies have been employed for this purpose. The former is an absorptive method depending on a change in permanent dipole moment during excitation, whereas the latter is based on small shifts in the frequency of scattered light due to interactions with vibrational states and requires a change in bond polarizability. Today,

infrared spectroscopy is almost always performed in a Fourier transform mode and is therefore (somewhat inappropriately) referred to as “FTIR.”
FTIR spectroscopy is the more commonly used of the two techniques at least partially due to the wider availability and lower expense of the instrumentation. The theory of IR absorption is briefly discussed in Chapter 4. The method has a number of advantages over CD including the ability to more easily monitor various states of matter (solid, liquid, gas, suspensions), an increased number of secondary structure sensitive signals, and the ability to perform experiments such as isotope exchange (see below), dichroism measurements,
P.536

and two-dimensional correlation analyses. Furthermore, in recent years it has begun to approach the sensitivity of CD.

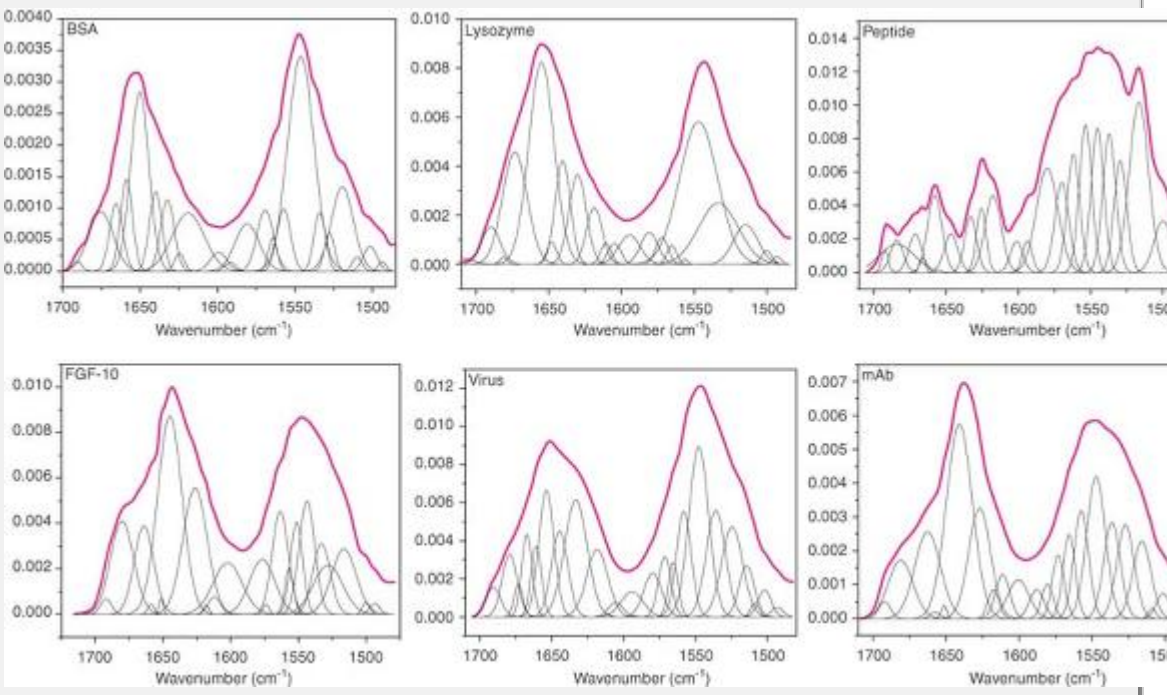

Fig. 21-16. Infrared spectra of proteins and peptides. The amide I region extends from 1700 to 1600 cm^{-1}. The spectra have been deconvoluted to show the relative contributions of the individual secondary structure types (and in the case of the peptide certain amino acid side chains) to the zero-order spectrum. Most commonly, the amide I region is used to estimate secondary structure content by assigning the origin of the individual bonds to different structural types (helix, β-sheet, turns, and disordered) which are ratioed to the total band area (see the text).

Infrared spectroscopy of proteins has been most frequently applied to the spectral signals known as the amide bands. There are many such vibrational absorption peaks with the most commonly examined the amide I, II, and III bands. The most frequently used band for secondary structure analysis in proteins is the amide I, which arises primarily from carbonyl stretching and to a lesser extent from NH wag. When proteins are examined, however, a very broad band is seen between 1600 and 1700^{-1} cm in this region (Fig. 21-16). A careful examination of this peak usually shows a number of small bumps and inflections suggesting the existence of underlying component signals. Several “band narrowing” procedures are available to deconvolute this broad peak into its constituents. These include derivative analysis as well as a procedure known as “Fourier self-deconvolution.” These component peaks are known to arise from

different types of protein secondary structure. Assignments are based on studies of model amino acid polymers as well as analysis of proteins of known (secondary) structure. Examples of assignments include α-helix (1650–1658 cm^{-1}), intramolecular β-sheet (1670–1680 and 1620–1640, multiple components), turns (1680–1700 cm^{-1}), and loops and disordered regions (1645–1655 cm^{-1}). In an observation of special importance to pharmaceutical scientists, aggregated proteins (manifesting intermolecular β-sheet) often produce distinct bands at either 1610 to 1620 or 1690 to 1695 cm^{-1}. Some amino acid side chains (Asn, Gln, Arg, Lys, Tyr, His) also produce signals in the amide I region but unless they are present at very high levels, their absorption bands are usually ignored. The major problem with the use of FTIR to analyze proteins in the past has been the strong absorption bands of water and water vapor in the region of the amide I band. For this reason, D_2O was often used as a solvent since its spectrum does not significantly overlap the amide I signals. Recent improved subtraction techniques, however, now permit the use of normal water for such studies. Thus, FTIR is now routinely used to estimate the relative secondary structure content of proteins with a precision and accuracy similar to that of CD.

As indicated previously, one of the advantages of FTIR is its ability to obtain spectra from samples in a variety of physical states. This is often accomplished through the use of different sampling techniques. For example, in addition to conventional transmittance geometry, spectra of both solids and liquids can also be obtained by attenuated total reflectance (ATR). In this technique, the sample is placed on a transparent plate of appropriate material which permits infrared radiation

to penetrate to a small extent into the sample allowing an absorption spectrum to be obtained. Solids can also be monitored by pressing them into a disc with a medium like potassium bromide (KBr) or illuminated from above and a reflective absorption spectrum obtained by a technique known as diffuse reflectance (DRIFT) spectroscopy. Infrared spectra can also be obtained through a microscope or a diamond-based cell, both of which permit very small areas to have their infrared spectra measured.

Nucleic acids can also be usefully examined by FTIR spectroscopy. In this case, the major signals of interest originate from the phosphate, base, and sugar moieties. All three groups are sensitive to nucleic acid conformation and to the binding of various ligands. Thus, B-form DNA can be distinguished from the other forms by the position of base carbonyl bands above 1700 cm^{-1} and the various phosphate-stretching vibrations in the 1000 to 1300 cm^{-1} region. Furthermore, the IR spectra of nucleic acids are quite sensitive to hydrogen bonding. Thus, the unwinding of DNA or RNA as measured in a thermal melting experiment can be monitored by FTIR spectroscopy.

Plasmid DNA is frequently used to deliver genes for gene therapy or as a vaccine. Often, the DNA is complexed to cationic lipids or positively charged polymers of various types. The cationic partners in such delivery complexes also produce distinct IR spectra. When the complexes are formed, IR signals from both components typically are altered as the various groups interact.[66,67] This permits complex formation and stability to be directly analyzed by this technique. For example, the CH_2 asymmetric stretching vibration of lipids can be used to measure the fluidization of their acyl chains in lipids bound to nucleic acids, providing a quantitative measure of their thermal stability and the effect of DNA upon lipid structure.[68]

Raman spectroscopy is also routinely used for all types of macromolecular-based systems.[64,65] This method has the advantages that the water bands are quite weak. This substantially decreases interference and many side chain vibrations are much better seen. Like FTIR, it can also be used to examine samples in multiple physical states. Its major disadvantage is that it is generally less sensitive than infrared spectroscopy. Two exceptions to this rule exist, however. If an absorption band can be directly excited, a spectral coupling process can produce dramatically enhanced vibrational signals from the specific chromophore (resonance Raman spectroscopy). In addition, if the molecule of interest is absorbed to certain types of materials such as silver, a much enhanced vibrational spectrum is again seen (surface-enhanced Raman spectroscopy or SERS). The mechanism of this enhancement is still debated. A major increase in use of Raman spectroscopy and its enhanced varieties is due to the

availability of tunable lasers of wide wavelength range and the potential to reduce the amount of material necessary for such measurements.

In the case of proteins, the most common application is the analysis of amide bands to obtain secondary structure information, analogues to applications of FTIR. It is more common with Raman spectroscopy, however, to use the amide III band because it is better resolved. General assignments are 1260 to 1305 cm^{-1} (α-helix), 1230 to 1245 cm^{-1} (β-sheet), 1258 to 1300 cm^{-1} (β-turn), and 1242 to 1255 cm^{-1} (disordered). Tyrosine side chains give well-resolved signals near 850 and 830 cm^{-1} and the ratio of these intensities has been used to analyze the relative exposure of phenolic side chains. Tryptophan (multiple peaks) and disulfide bands (500–550 cm^{-1}) also produce conformationally sensitive signals. Thus, secondary and tertiary structure changes can be simultaneously examined by Raman spectroscopy, a particularly attractive aspect of the technique.

Raman spectroscopy of nucleic acids has also frequently been used to explore nucleic acid structure, perhaps more than infrared absorption spectroscopy. The two major forms of DNA manifest differences in the 800 cm^{-1} region (phosphodiester antisymmetric stretching) with this signal present in the A form but absent in the B form. The left-handed Z-form produces a unique peak near 625 cm^{-1}(shifted from 675 cm^{-1} in the A and B forms). A variety of other conformationally sensitive signals are available in other regions of the spectrum. Raman spectroscopy has also been widely used to characterize viruses in which distinct signals from both protein and nucleic acid components can be easily resolved.65 Furthermore, the interaction between these two components can be analyzed as well as various aspects of viral structure and stability. Few studies of this type have yet been performed with IR spectroscopy.

Scattering,69,70,71,72,73***Hydrodynamic,***74,75 ***and Calorimetric***76,77,78,79,80,81 ***Techniques***39

The theory behind many of the methods described in this section is discussed in Chapter 16 and should be reviewed accordingly. Specific applications to biopharmaceuticals and topics not reviewed previously will be focused on here.

Light scattering is an extremely useful technique for analyzing the size and shape of biomolecules. For macromolecules which are much smaller (<*λ*/50) than the wavelength (*λ*) of light used in a scattering experiment, it can be shown that:

$$\frac{\mathrm{I}_\theta}{\mathrm{I}_\mathrm{o}} = \frac{2\pi \mathrm{n}_o^2\left(\frac{\mathrm{dn}}{\mathrm{dc}}\right)^2}{N\lambda^4 \mathrm{r}^2}(1+\cos^2\theta)\mathrm{M}_\mathrm{wC} = \mathrm{KM}_\mathrm{wC}$$

where I_θ is the intensity of the scattered light at some angle θ, I_o is the incident intensity, n_o is the refractive index of the solvent, dn/dc is the variation in refractive index of the solution with variation in concentration of the scattered (the refractive index increment), N is Avogadro's number, M_w is the weight average molecular weight, and$c$ the concentration. Thus, in this case the intensity of scattered light is proportional to the molecular weight. The concentration dependence of this expression can also be used to calculate virial coefficients, which can be used to characterize the interaction between molecules. In the case of large molecules, which

possess multiple scattering centers within themselves, it is found that:

$$\frac{K_\mathrm{c}}{R_\mathrm{o}} = \left[1 + \frac{16\pi^2 {R_\mathrm{g}}^2}{3\lambda^2}\sin^2\frac{\theta}{2}\right]\left[\frac{1}{M_\mathrm{w}} + 2Bc\right]$$

where $R_\theta = \frac{r^2}{1+\cos^2\theta}\frac{I_\theta}{I_o}$ (the "Raleigh ratio"), and

$$K = \frac{2\pi {n_o}^2 \left(\frac{dn}{dc}\right)^2}{N\lambda^4}, \quad M_w = \frac{\sum n_i {m_i}^2}{\sum n_i m_i}$$

where B is the second virial coefficient and R_G is the radius of gyration which is defined as $\Sigma m_i r_i^2/(m_i)$, where m_i is the mass of the ith element at distance r_i from the center of mass of the scattering particle. Inspection of this somewhat complex expression, however, shows that a plot of $\frac{Kc}{R_\theta}$ vs $\sin^2\frac{\theta}{2} + Kc$ yields an intercept equal to the molecular weight (M_w) and a slope proportion to M and R_G. The physical meaning of R_G is not obvious but it has a fairly simple relationship to various shapes. For example, for a sphere, $R_G = R\sqrt{3/5}$ where R is the radius of a sphere, a rod, $R_G = L\sqrt{1/12}$ where L is the length of a rod and $R_G = \frac{\sqrt{\overline{h}^2}}{\sqrt{6}}$ for a random coil, where $\overline{h}^2$ = mean square end-to-end distance. More generally, shapes can be modeled as prolate or oblate ellipsoids. For example, for a prolate ellipsoid, $\overline{h}^2$ where 2a, 2a, and Y2$a$are the axes of the ellipsoid. Given these types of relationships, the shape of a particle can be estimated from the ratio of the value of the observed R_G to the calculated R_G for a sphere. Globular proteins typically give a value near 1, while more elongated molecules like myosin and DNA produce values greater than 10. If greatly elongated molecules like DNA give observed values of the R_G much less than those calculated for an equivalent rigid rod, this is direct evidence for their flexibility. Another way to think about scattering by large particles is in terms of their turbidity ($\tau = \frac{1}{\ell}\ln\frac{I_o}{I}$ and $\tau = H_c M_w$). This is discussed in more detail in Chapter 16. Although this approach is less sensitive than studying scattering at other angles, it can be simply obtained in an absorption spectrometer since it is simply related to decreases in transmittance. Turbidity measurements are widely used in biopharmaceutics for this reason and have been especially widely applied in kinetic studies of macromolecular aggregation (see below).

A second type of light scattering experiment that has been widely used in biopharmaceutics is "dynamic" or "quasielastic" light scattering (DLS or QELS). This involves an analysis of the fluctuations in intensity

of scattered light due to Brownian motion of the scatterers. This analysis is in terms of what is known as an "autocorrelation function," $G(\tau)$ which is defined as $\langle I(t) \times I(t+\tau) \rangle$ where $I(t)$ is the intensity of scattered light at time t and $I(t+\tau)$ is the intensity at some short time (τ) later. Inspection of this function finds that if the intensity remains high as τ is increased, its value will be high. If τ increases, however, and the value of $I(t+\tau)$ changes rapidly, it will quickly time average to zero. Thus, the value of the autocorrelation function falls toward zero more rapidly for smaller, faster moving molecules (Fig. 21-17). This decrease is exponential in form and is given by:

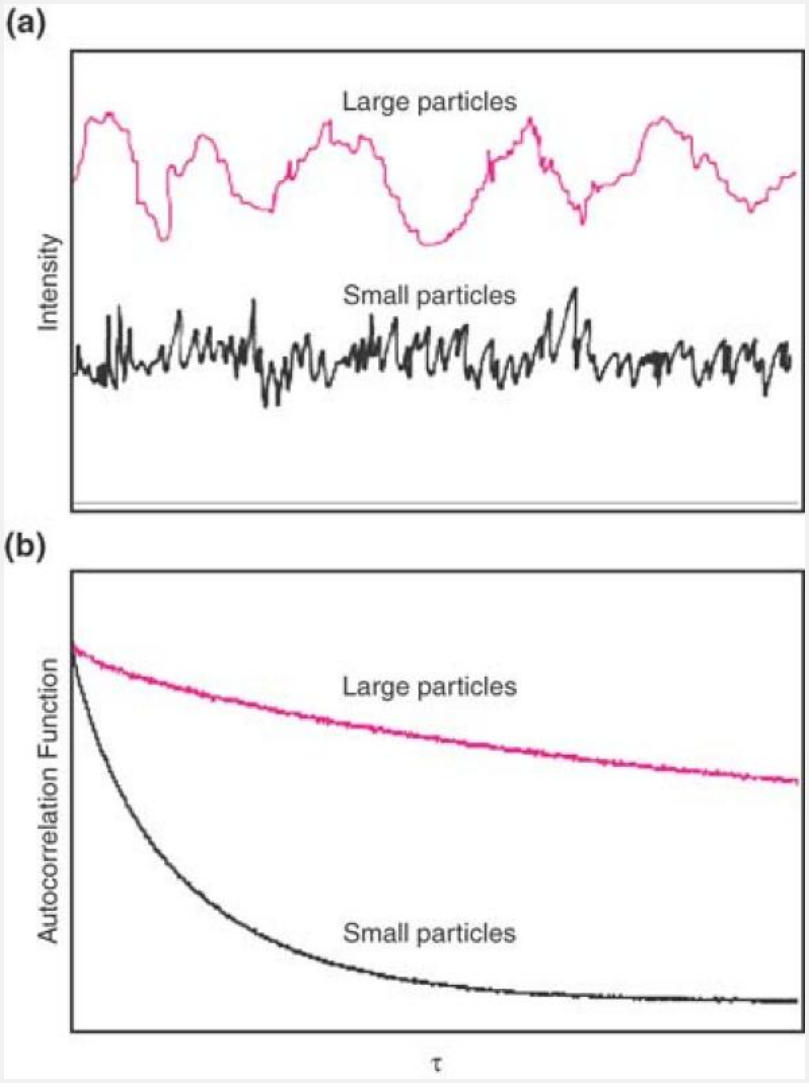

Fig. 21-17. Dynamic light scattering. The size of particles can be determined by measuring the fluctuations in the intensity of their scattered light (*a*). Because of their Brownian motion, small particles produce more rapid changes in scattered intensity than larger ones (*b*). If one measures the change in intensity at very short time intervals, the intensity changes less quickly for light scattered by the larger particles and we say that these intensity changes are more "auto-correlated." Analysis of autocorrelation functions (see text) allows the size (more specifically the hydrodynamic radius or diameter) of the scattering particle to be determined.

$$G(\tau) = A \exp(-2DQ^2\tau) + B$$

where D is the diffusion coefficient of the scattering molecule, B is a constant, and Q is given by:

$$Q = \frac{4\pi n}{\lambda} \sin \frac{\theta}{2}$$

where n is the refractive index and θ the angle at which the scattered light is observed. Thus, the angular dependence is small for smaller molecules ($d < \lambda/10$) but larger at lower angles for larger molecules.D can be related to R (the Stoke's radius of the scatterer) through the Stokes–Einstein equation in the following convenient form (see equation 17-7):

$$D = \frac{kT}{6\pi\eta R}$$

P.539

where η is the solution viscosity. Thus, the size of a large molecule like a protein or nucleic acid can be directly measured by this method. If the solution is homogenous, an accurate size is easily obtained. If it is not, two approaches can be taken. In the first, the data are fit to a function of the form ln [$g(\tau)$ - k] to yield two parameters, a weight average mean diameter and a polydispersity parameter (a measure of the width of the size distribution). This is known as the method of cumulants. In the second, the data are fit to a sum of exponential functions to yield a multimodal distribution of sizes. If the different species differ by more than a factor of two in diameter, they can usually be resolved if distinct populations are present. Analysis can be in terms of weight, number and intensity distributions, with the number distribution usually the most intuitively useful. If there are large, internal fluctuations within a macromolecule as might be seen by a large, flexible molecule like DNA, DLS may also be able to detect and resolve these motions as contributions to autocorrelation functions. Dynamic light scattering has become an increasingly important tool to the biopharmaeutical scientist as highly convenient commercial instrumentation has become increasingly available and changes in size and aggregation state are recognized as important degradation pathways for biotechnology products.

A number of other methods often prove useful in the analysis of macromolecular size and shape. Osmotic pressure and viscosity measurements have previously been discussed in earlier chapters and will not be further considered here. More frequently used in the last few years, however, is analytical ultracentrifugation. This is primarily due to the introduction of modern instrumentation to perform such studies. Recall that in a velocity sedimentation experiment, the experimentally measured sedimentation coefficient (s, the velocity of the sedimenting particles divided by the unit centrifugal acceleration) is directly proportional to the molecular weight, allowing this quantity to be determined if the diffusion coefficient (D) and partial specific volume ([v with bar above]) are known ($M = \frac{RTs}{D(1-\bar{v}\rho)}$). If a mixture of macromolecules has multiple components which differ significantly in S, they can sometimes be resolved by this technique (**Fig. 21-18**). It is possible to perform sedimentation analysis in gradients of substances such as sucrose and cesium chloride. This has the advantage that differences in densities of particles are exploited. Therefore, mixtures containing proteins, nucleic acids, and viruses which differ in density can be separated and analyzed, and to some extent at a preparative scale.

If instead of measuring the velocity of sedimenting particles or using solute gradients, one spins a solution of macromolecules into an equilibrium gradient, one can also calculate molecular weight from the resultant distribution of mass. This is described by:

$$M_w = \frac{2RT}{\omega^2(1-\bar{v}\rho)} \frac{d \log c}{d(x^2)}$$

where ω^2 is the angular velocity of the rotor at equilibrium and d log$c$/d(x^2) describes the gradient of concentration as a function of the distance from the center of the rotor. Thus, the slope of a plot of d log$c$ versus dx^2 allows the molecular weight to be determined. Note that the need to know the diffusion coefficient has disappeared and therefore one can obtain an absolute estimate of the molecular weight from an equilibrium sedimentation study. A quite large molecular weight range (less than 100 to more than 10 million Daltons) can be characterized by this method. A very powerful application of equilibrium sedimentation involves the analysis of associating or dissociating systems. This is accomplished by fitting the data to various models of such behavior. The sensitivity and accuracy of this method permits both stoichiometries and equilibrium constants of associating and dissociating macromolecules to be obtained.

Two versions of microcalorimetry are also widely used in the analysis of biopharmaceuticals. As described in Chapter 2, differential scanning calorimetry (DSC) measures the excess heat capacity of a molecule as a function of temperature. If there is an absorption or release of heat due to a structural change in a macromolecular system (an endothermic or exothermic transition), a peak is usually seen in a DSC experiment. A plot of Cp versus T is known as a thermogram. If the process is reversible, the area under such a curve corresponds to the enthalpy (ΔH) of the change in state. Structural changes in biopharmaceuticals are often detected by this technique (Fig. 21-19). Protein unfolding, the melting of nucleic acids, and phase changes in lipid bilayers are all routinely studied by this method. If transitions are not reversible, the temperature at which the peak of the transition occurs (the "T_m" or melting temperature) is used as a measure of thermal stability. In many cases, DSC thermograms can be quite complex. This can occur for a variety of reasons. For example, samples such as membranes or viral particles which contain multiple components can correspondingly manifest several transitions. Furthermore, individual structural domains within individual proteins can also often be resolved as isolated or overlapping thermal events. For example, the multiple structural domains present in immunoglobulins usually produce multiple peaks in the thermograms of these molecules (Fig. 21-19). When ligands bind to macromolecules, they often perturb their stability and thus can be detected as a change in T_m. Furthermore, protein aggregation can sometimes be seen as exothermic transitions, in contrast to the endothermic events seen as bonds are broken in other processes. Because it is not dependent on the presence of specific chromophores, is reasonably sensitive (sample concentrations as low as 10 μg/mL have been employed) and is now available in a high throughput (HTP) (autosampling) mode, DSC is widely used as a routine tool in the characterization and formulation of biopharmaceuticals (see below).

A second calorimetry method of importance to the pharmaceutical analysis of biopolymers is isothermal titration calorimetry (ITC). In such experiments, small amounts of one component are incrementally introduced to a second. A common application is to study the interaction of a small molecule (ligand) such as an enzyme effector or excipient

stabilizer to a protein or nucleic acid. The small heats produced during the binding interaction can be plotted as a function of the molar ratio of ligand to receptor and the data fit to various binding models (see Chapter 11). If a good fit can be obtained, quantitative analysis of such data can yield the binding constant (free energy), enthalpy, and entropy of the interaction as well as the stoichiometry of the binding event. Thus, this approach has been widely used to study the interaction of plasmids with their delivery partners.[82,83] The heats of dilution produced as a consequence of the titration process must be subtracted from the experimental binding heats. The sample requirements for ITC analysis are modest (0.1–1 mg) and the method has also been adopted for HTP applications.

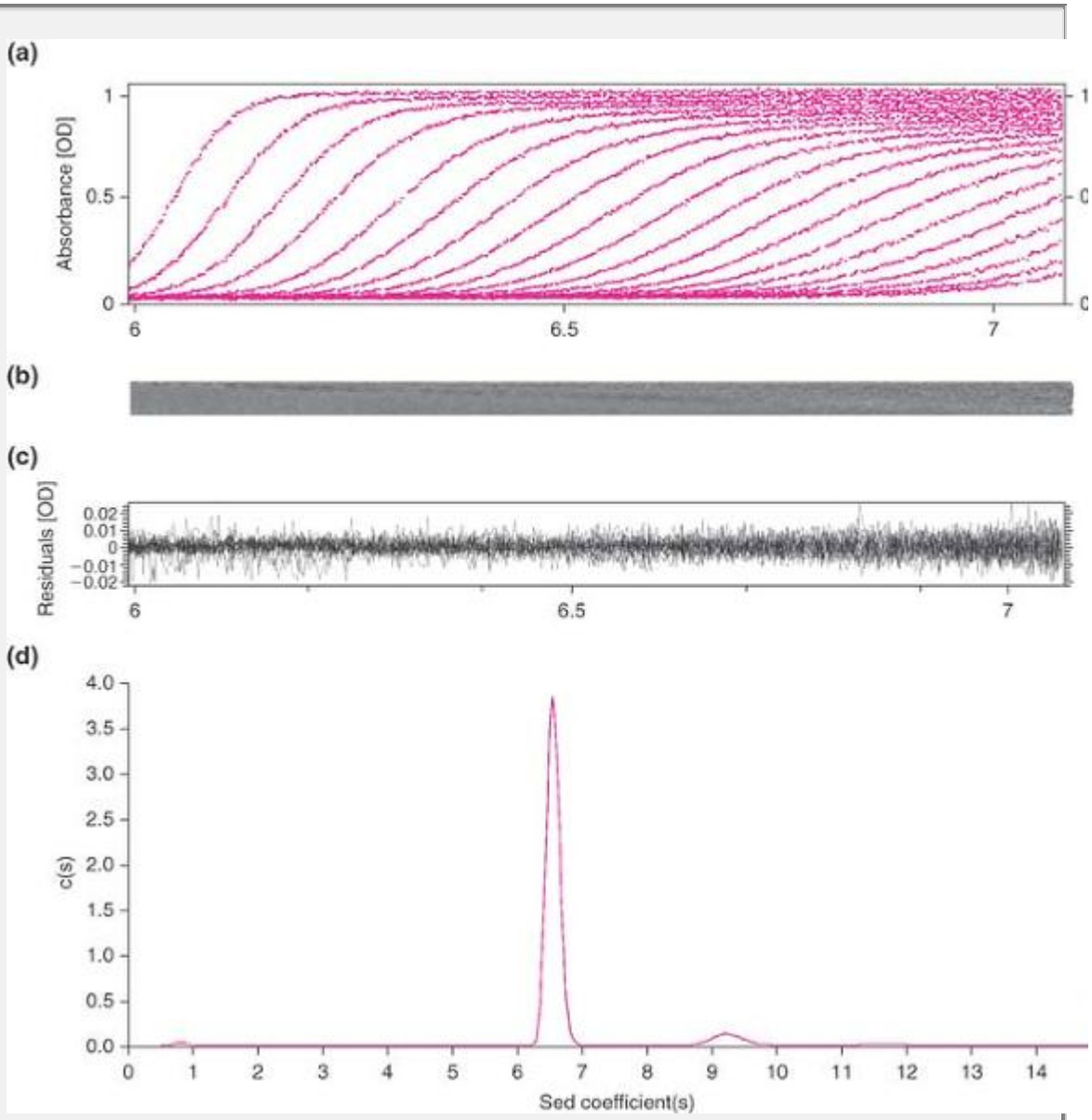

Fig. 21-18. (*a*) Sedimentation velocity profiles of a monoclonal antibody. The data was acquired from the absorbance at 280 nm. (*b*) and (*c*) are, respectively, the residue bitmap and residues of the best data fit. The fitting was done with the program Sedfit at a confidence level of 0.68 with sedimentation coefficients seen from 0.5 ~ 15, the best fit friction ratio at 1.491 and partial specific volume at 0.728. (*d*) Continuous sedimentation coefficient distribution of the monoclonal antibody finds the sedimentation coefficient of the antibody monomer of the antibody at 6.53 and the dimer at 9.20.

Analysis of Macromolecular Dynamics

The previous techniques discussed for use in macromolecular analysis are in general time averaged methods in that they see an averaged property that is smeared out over the time of the measurements which usually take many seconds to hours. Molecular systems like proteins, nucleic acids, and viral particles, however, display a wide variety of much faster motions that play an important role in their structure, function, and stability.[84,85] In the case of proteins, these internal

motions range from the rotations and flexing of individual side chains through movements and telescoping of regions of secondary structure to large-scale motions of entire domains. The presence of such motions means that probably the best picture of a protein molecule in solution is that of a large, Boltzman-like distribution of rapidly interconverting conformational states, with the true native state of any protein best described by such a distribution (see Fig. 21-21, later in this chapter). The importance of this view of protein structure is just beginning to be recognized in the world of biopharmaceuticals because of its relationship to physical and chemical stability. Nucleic acids also display marked internal motions, the most common characterized as “breathing modes.” This involves the rapid breaking and remaking of hydrogen bonds between the bases as well as changes in the stacking interactions between the bases. Although these fluctuations are quite small in large DNA molecules, they can be quite significant in regions of stress or at the ends of duplexes or at ss/ds junctions. Similarly, lipid bilayers are also subject to significant thermal motions that play a key role in their structure and functional properties.

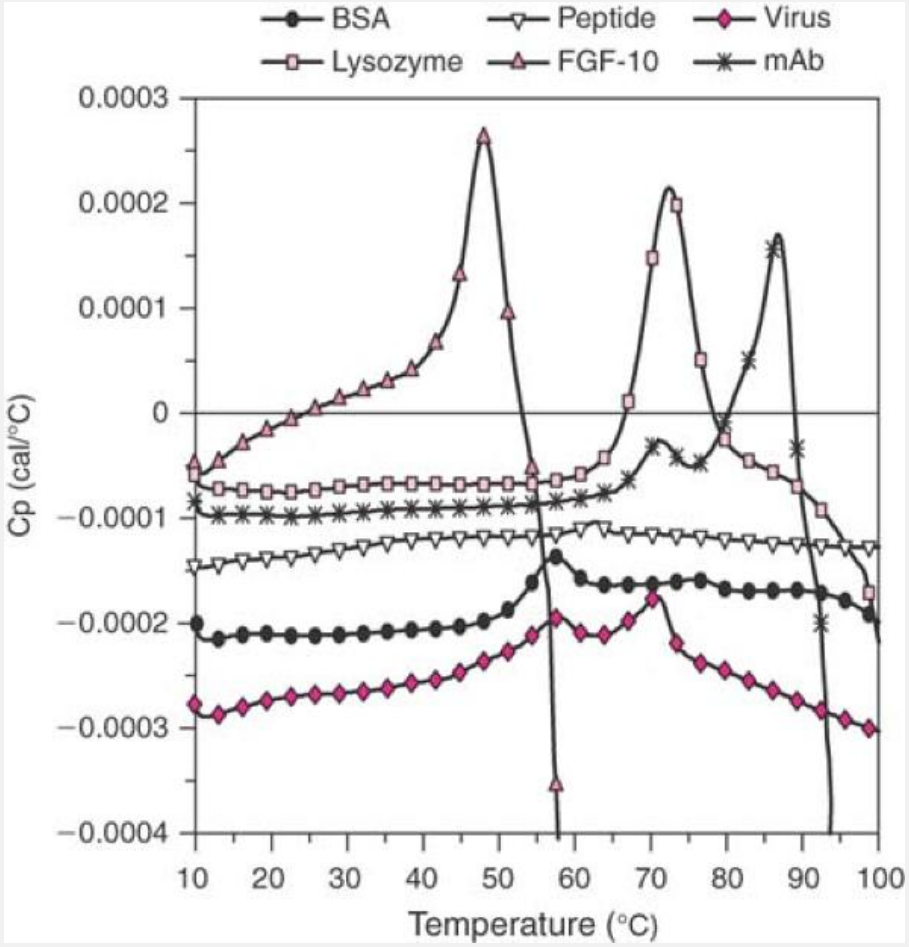

Fig. 21-19. Differential scanning calorimetry (DSC) of the representative biopharmaceuticals. The sudden decrease in heat capacity seen with FGF-10 and the IgG is due to their aggregation and precipitation as the temperature is raised. Note the multiple transitions seen with the IgG and BSA, both of which are multidomain proteins (see **Fig. 21-1**). The multicomponent nature of the viral particle also produces at least two transitions.

A number of experimental methods that can detect these types of rapid motions have already been described. These additionally include isotope exchange measurements as detected by NMR, FTIR, and mass spectrometry as well as molecular dynamic simulations. The availability of an array of novel multidimensional NMR methods also provides the ability to directly monitor the dynamic behavior of individual residues. The results of isotope-exchange studies (employing deuterons) usually reveal at least three different classes of exchangeable protons from peptide bonds: (*a*) a rapidly exchanging group that exchanges too fast to be detected (i.e., in less than a few seconds), (*b*) a class that exchange over seconds to many hours, and (*c*) a small number of buried protons that do not exchange over the lifetime of the experiments. The binding of ligands or alterations in protein/protein interactions typically produces changes in the relative number of each class of exchangeable protons. This then provides one picture of the dynamic aspects of protein structure in terms of the accessibility of the peptide backbone to solvent water.

A wide variety of other methods are also available to probe similar and different aspects of protein dynamic behavior. As discussed above, several fluorescence-based methods provide alternative pictures of internal protein motions. Instead of the use of proton exchange, the quenching of tryptophan (and to a lesser extent tyrosine) residues can be used to study protein motions that permit the diffusion of various solutes into different protein regions. Similarly, as mentioned previously, cations of various sizes can be used in the same manner by measuring shifts in the derivative absorption peaks of Trp, Tyr, and Phe due to cation/pi interactions. The increased number and type of residues in the latter approach offers several advantages. In a new method, the slopes in the shifts of these same derivative absorption peaks with temperature can be used as a measure of protein motions.86 This is based on the well-understood temperature dependence of the dielectric constant of water and solvent penetration into protein interiors with highly buried aromatic side chains producing little or no temperature-dependent slopes in contrast to more exposed ones.

Time-resolved fluorescence anisotropy methods can be used to sample very rapid motions in the picosecond to nanosecond range of times. In this technique, polarized photons are used to excite fluorophores and their depolarization upon emission is used to characterize the motion of individual molecular groups in terms of their rotational correlation times. Because the fluorescence lifetimes of indole are so short, only rapid motions can be seen in this case. But if extrinsic fluorophores with longer lifetimes are either covalently attached or noncovalently bound to specific sites on a protein or nucleic acid, larger scale motions can be sampled. For example, if a long lifetime fluorophore is placed in the antigen-binding site of an antibody or is attached to a cysteine residue at a defined location, motions such as the flexing of the arms of a Y-shaped antibody can be measured. Another dynamics-sensitive fluorescence-based technique is red-edge excitation in which slow dipole relaxation and photoselection are on the same (or longer) timescale than fluorophore lifetimes. Because these processes are solvent dependent, they can be related to the rigidity of the local environment. If this local matrix is not altered, the emission wavelength of a subensemble present may be uniquely excited and be of lower energy than the mean distribution. Thus, the fluorescence emission spectrum will be excitation dependent and shifted to longer wavelength.

Two relatively simple techniques can also be used to measure the expansivity and contractibility of proteins, both parameters related to their dynamic behavior. In pressure-perturbation isothermal titration calorimetry, the heat emitted or absorbed when pulses of pressure are applied differentially to a sample and reference is measured. This heat difference can be used to determine the coefficient of thermal expansion of the partial volume of the target macromolecule. Such studies also permit the accessible surface area and solvation to be obtained. In complementary measurements, ultrasonic spectroscopy can be used to obtain the adiabatic and isothermal compressibility of a sample of any type. High-frequency sound waves are sensitive to intramolecular interactions because they produce compressions (and subsequent relaxations) of highly structured polymeric systems. By measuring the speed of sound

through such materials, the attenuation produced by the pressure-induced compressions and decompressions can be related to the presence of cavities in macromolecular interiors. This is, in turn, related to fluctuation in volumes and their coupling to the local solvent and thus protein dynamic behavior (Fig. 21-20).

A large number of other techniques are available to probe the dynamics of higher molecular weight systems. These include neutron diffraction, single molecule fluorescence spectroscopy, three pulse photon echo peak shift spectroscopy, ultrafast two-dimensional vibration echo, and correlation spectroscopy among others. We will not discuss them here, but the interested student needs to be aware that this is a rapidly expanding field with new approaches routinely becoming available.

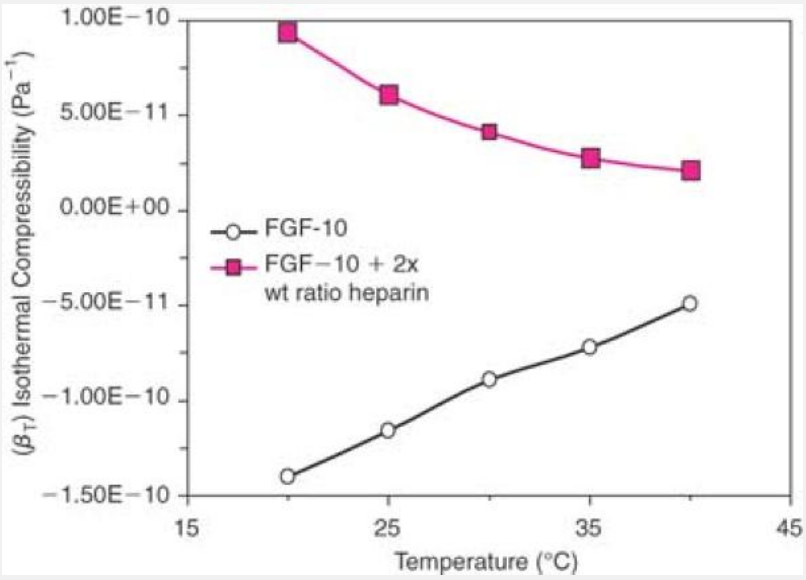

Fig. 21-20. Some idea of the effect of a ligand on the dynamics of the protein behavior can be obtained by measuring its compressibility. This is done by measuring the attenuation of sound when it is passed through a protein-containing solution (ultrasonic spectroscopy). As the sound-induced compression of the solvent squeezes a protein, a certain amount of energy is lost as the protein is compressed. This loss can be used to estimate the compressibility. In this example, when heparin is added to fibroblast growth factor-10 (FGF-10) (square), the protein appears to become more compressible, that is, its range of dynamic motion increases. The difference in compressibility is decreased, however, as the temperature is raised. See reference 84 for further discussion.

Breathing modes of nucleic acids can also be measured by many of the above methods. A number of unique methods are available as well. For example, chemical probes such as formaldehyde or dimethylsulfate that specifically react with single-stranded sequences can be used to measure the fluctuations in duplexes that are responsible for their reactivity. It is also possible to replace adenine bases in DNA and RNA with 2-aminopurine, which possesses unique CD, and fluorescence spectral properties that can be used to sense local dynamic behavior. A wide variety of fluorescence and electron spin resonance probes can be used to study membrane dynamics along with NMR and variations on some of the methods described above.

What exactly is the utility of the many dynamics-sensitive methods briefly indicated here? Although such studies are still in their infancy, it is clear that an intimate relationship exists between the dynamics of biopharmaceutical systems and their stability. It was initially thought that this correlation might be a simple one in which increased rigidity (reduced dynamic motions) was related in increased stability.

Although this relationship has sometimes been observed, it has also been found that local decreases in stability can be observed upon ligand binding and macromolecule/macromolecule interactions. This is probably due to increases in rigidity and stability in one region of a molecule being relieved by decreases in stability (and increases in dynamic behavior) in other parts. It is clear, however, that the role of internal dynamics is becoming better understood in macromolecular system and that these phenomena will play an increasingly important role in the stabilization and formulation of biopharmaceuticals.

Preformulation87’88’89’90

The insertion of a molecule into a chemical system in which it possesses sufficient solubility, stability, and deliverability such that it can be used as a drug or vaccine is commonly referred to as “formulation.” The final form of this system containing the molecule itself as well as its accompanying excipients used to achieve these acceptable properties is also described as its formulation (noun). The formulation of biomolecules follows a process generally similar to that used for smaller molecules except that the physical nature of these much larger molecules necessitates the use of many different experiment methods (see above) as well a variety of other considerations based on their unique properties. It is conventional to consider the degradation of biomolecules as either physical or chemical in nature. In general, by degradation we mean change in structure. This may or may not be accompanied by a loss in biological activity as described below. This initial analysis of a macromolecular system prior to the preparation of the formulation is referred to as “preformulation.”

P.543

Physical Degradation

Physical degradation is characterized by changes in the noncovalent interactions within and between biomolecules. The relationship between physical and chemical degradation will be considered in a later section. Physical degradation is usually discussed in terms of a catalogue of the various types of noncovalent interactions. Here, however, these phenomena will be considered in a somewhat different manner. Imagine a protein in solution of average stability at moderate concentration (0.01–10 mg/mL) at a fixed, near neutral pH (5-8). What happens to the macromolecule when we raise and lower the temperature starting under ambient conditions (15°C–30°C)? The effect of temperature will be considered from two different perspectives: changes in the structure of an individual protein and alterations in the distribution of the microstates of a population of such molecules (Fig. 21-21). Similar comments are applicable to nucleic acids and other macromolecular systems. Temperature is chosen as the “stressing” variable here because of its general nature and critical role in the storage stability of biopharmaceuticals. The reader is reminded that the primary effect of temperature is on the thermal motion of the solvent (the water molecules) and the internal motions of the various molecular entities within the protein. Increasing the temperature, of course, increases the rate and magnitude of such motions while lowering it does the opposite.

As the temperature gradually begins to rise the interior motions of the protein will begin to gradually increase. Thus, the local motion of side chains, larger scale motions of elements of secondary structure, and the translational movement of the protein all begin to gradually increase in magnitude. At least at first, these effects on the structure of a protein are usually difficult to detect although they can often be seen in continuous changes in parameters like UV-absorption derivation peak positions. Most importantly, they generally have little immediate, obvious effect on the structure and stability of most proteins. This also results, however, in an alteration in the distribution of microstates with a shift to higher energy. As the temperature is further increased, however, these increases in internal motions can lead to a significant weakening of many forces that stabilize protein structure such as hydrogen bonding, electrostatic, and Van der Waals interactions.91 Note that the strength of apolar (hydrophobic) interactions tends to increase as the temperature is raised because of favorable entropic ($-T\Delta S$) and heat capacity (ΔCp) effects. In some proteins, this can lead to actual conformational alterations with the distribution of states splitting into two or more peaks. This does not necessarily imply that the protein

must unfold (denature) under these conditions, but rather that the weakening of key intramolecular interactions can produce local alterations in structure. This could be due to unfoldinglike changes in a particular region of the protein, a change in structure of one or more domains in a multidomain protein, or a consequent dissociation of a subunit-containing protein, among other possibilities. Such alterations could result in a change in a protein's biological activity or its immunogenicity. It could also lead to aggregation of the protein, especially if an apolar region becomes exposed. The type of conformational changes indicated may or may not be detected by methods like CD, intrinsic or ANS fluorescence, or DSC, depending on their magnitude and the exact nature of the structural change. As mentioned previously, one special form of these types of structural changes produces an important class of altered protein configurations known as MG states.92

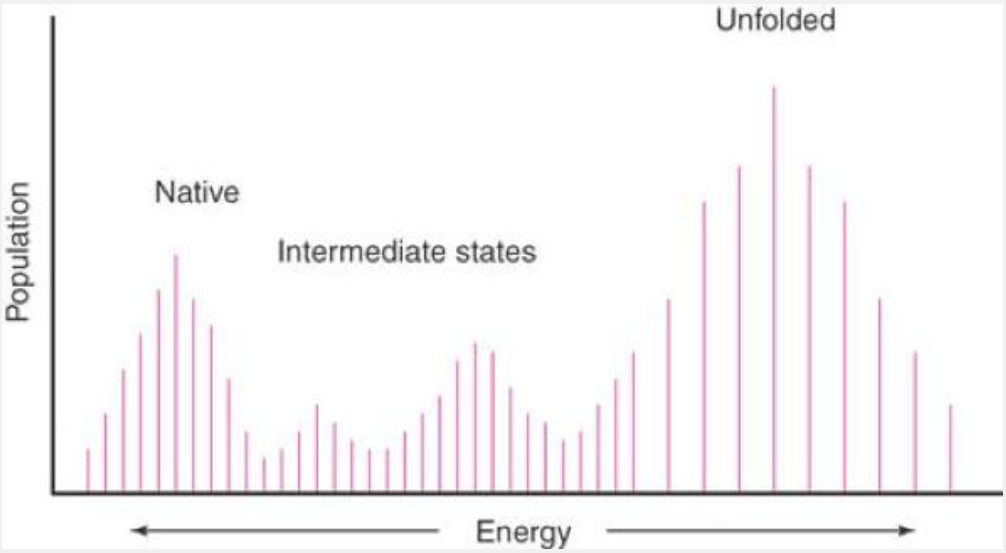

Fig. 21-21. The behavior of proteins and other macromolecular-based systems can be described in terms of a distribution of different structural states, each possessing certain energies. As the environment of the macromolecules changes as a result of alterations in solution variables such as pH and temperature, the distribution itself will alter. For example, at lower temperatures, lower energy state may be more populated (i.e., "native states"). At higher temperatures, as bonds are broken and the structure is altered, the population of these states (i.e., "unfolded states") will increase. The existence of intermediate states is equally important due to their unique, individual behavior.

MG states have several distinguishing characteristics. They display a dramatic decrease in tertiary structure as seen by exposure of their aromatic side chains as monitored by intrinsic fluorescence, ultraviolet absorption, near UV CD or related techniques. In contrast, their secondary structure remains substantially intact as seen by far UV CD, infrared, or Raman spectroscopies. Thus, when proteins that display this state are heated, the contacts between secondary structure elements and other distant contacts within the polypeptide chain are broken prior to major alterations of secondary structure. One consequence of this is that dyes such as ANS usually bind to MG-states with a dramatic increase in fluorescence, aiding in their identification. The reason MG states are so important to the pharmaceutical scientist is their tendency to aggregate. It is now generally thought that many cases of aggregation are due to the population of such states (Fig. 21-2). They are also commonly seen in proteins at low pH and high salt concentration, but their transient presence is probably responsible for many if not most cases of protein aggregation. If a protein continues to be heated, a much more comprehensive disruption of structure may take place. Although these so called "unfolded" or "denatured" states usually still contain

some structure (especially in thermally induced unfolding), a loss in biological activity is typically produced. When proteins begin to unfold, they often interact

with themselves to form intermolecular β-structure. These aggregative states can usually be identified by unique FTIR signals and dye binding as mentioned above and appear to be involved in a variety of disease states such as Alzheimer and Parkinson disease in which precipitated protein is present in vivo. Complete unfolding does not necessarily (and probably rarely) produce aggregation. Although this was once thought to be the case, passage through MG states is more likely responsible for most examples of commonly observed thermal aggregation. Although thermally induced aggregation is usually (but not always) irreversible, use of denaturing agents such as urea, guanidinium hydrochloride, or a chaotropic salt (e.g., $LiClO_4$, $NaSCN_4$) can often be used to produce reversibly and more extensively unfolded protein. Frequently, data obtained from such experiments can be modeled as a simple, reversible, two state unfolding transition in which intermediates do not play a significant role[93,94] (Fig. 21-22):

$$\mathrm{N} \geq \mathrm{U}$$

where N and U refer to native and unfolded states, respectively. The fraction unfolded (fu) as induced by temperature or unfolding agent is then given by

$$f_u = \frac{X_N - X}{X_N - X_U}$$

where X is the experimental value determined from a method such as CD, fluorescence, and so forth for the native (X_N), unfolded (X_u), and fractionally unfolded (X) state. The equilibrium constant for unfolding (K_u) is then given by

$$K_u = \frac{[U]}{[N]} = \frac{f_u}{1 - f_u}$$

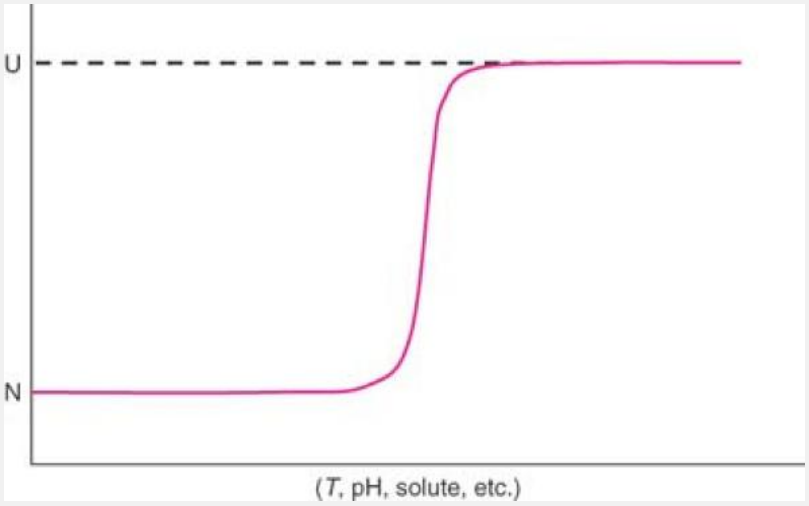

Fig. 21-22. A generic "unfolding curve" for a simple two state system. If only two states, such as a native and unfolded form, are detectably present, a very sharp, highly cooperative transition is usually seen. Such transitions can be induced in macromolecular systems by a wide variety of different variables including temperature, pH, and the presence of solutes such as urea and guanidine hydrochloride. The midpoint of such curves when temperature is used as the perturbing variable is known as the melting temperature or the T_m.

and the free energy (ΔG_u) by

$$\Delta G_u = RT \ln K_u$$

The relationship $\Delta G_u = \Delta H_u - T\Delta S_u$ and the Van't Hoff equation permit estimates of the associated thermodynamic parameters for the unfolding process.
If a denaturant is used[93,94,95]:

$$\Delta G_u = \Delta G_{H_2O} - m[\text{denaturant}]$$

where ΔG_{H2O} is the extrapolated valve of the free energy at zero denaturant concentration and m (the dependence of ΔG_u on denaturant concentration) is a parameter related to the amount of protein surface area that becomes solvent exposed upon the induced unfolding. Thus, at least in these special circumstances, a fairly comprehensive quantitative picture of the unfolding process can be obtained.
Let us return to more moderate temperature conditions. It is also possible for proteins to aggregate without any conformational change in their structure. The highly amphipathic nature of protein surfaces (i.e., they possess both polar and apolar regions) means that they have a significant potential to interact with themselves. Such interactions are usually temperature dependent with increased aggregation seen as the temperature is lowered. As discussed above, however, increases in temperature can also lead to aggregation through conformational changes or be due to the increasing strength of apolar interactions at higher temperatures. A variety of other forms of environmental stress important to the pharmaceutical scientist such as shaking and freeze/thaw events can also lead to protein conformational changes and/or aggregation although these phenomena are less well understood despite their pharmaceutical relevance.
Low temperature can also promote the unfolding of proteins due to the reversed temperature dependence of the hydrophobic effect and the large heat capacity (and surface area) changes, which are usually associated with protein conformational alterations. Maximum destabilization is often seen at temperatures below freezing but destabilization per se can frequently be a factor in the behavior of proteins at quite moderate temperatures. Another form of physical degradation also arises from the amphipathic nature of protein surfaces. Proteins in solution must be resident in some type of container. This presents a variety of different types of surfaces with which a protein can interact. In fact, the plastic and glass containers commonly used to store proteins and other biopharmaceuticals may themselves possess some charged or apolar characteristics, enhancing the potential for protein/surface interactions. The air/water interface produced during shaking can also be considered an example of such a surface. Rubber stoppers and various types of syringes are frequently siliconized to reduce such interactions but often with only limited success. In fact, at very low protein concentration (<10 μg/mL), a substantial portion of most proteins may be adsorbed to container surfaces. As proteins remain adsorbed to surfaces, they may undergo conformational changes that optimize their interaction with the

surfaces, leading to additional problems. Potential solutions to such problems are described below.
The effect of environmental factors other than temperature can also play an important role in the physical degradation of proteins. The highly charged nature of protein surfaces (and to a much lesser extent their interiors) makes them very pH sensitive. Important approximate pK_a's are Asp (3.0), Glu (4.2), Lys (10.0), Arg (12.5), His (6.0), Tyr (10.0), and Lys (9.1). These valves can vary quite significantly (by several pH units) in proteins because of local environmental effects.[96,97] Rough estimates of the total change on a protein can be made from these values or from average valves of the individual residues based on actual measurement of a large number of proteins. The charge density of a protein (P_c) can be crudely estimated from[39]:

$$P_c = \frac{(pI - pH)}{M_w}$$

where the pI is the isoelectric point of the protein (the pH at which the charges sum to neutrality) and M_w is the molecular weight of the protein. Proteins tend to display their minimum solubility near their pI although many exceptions to this rule exist because of the potential for specific interactions among protein molecules.
Charged residues often provide key interactions in the stabilization of protein structure. In addition to direct interactions between oppositely charged side chains (sometimes called salt bridges or ionic

interactions), ion–dipole and cation–pi interactions are commonly seen in proteins. Thus, as pH is varied, any of these (and many more) types of interactions can be altered leading to structural changes and changes in the stability of proteins.

The presence of salt has a major effect on the electrostatic behavior of proteins. At low salt concentrations (<0.2 M for a simple salt like NaCl), this is successfully modeled by a simple charge-shielding model as originally described by Debye–Hückel theory. This also works well for the interactions of ions like Mg^{2+} with the phosphate backbone of DNA. As salt concentrations are raised, however, a variety of other phenomena are observed. In the case of proteins, solubility can be dramatically decreased to the extent that certain ions can produce precipitation from solution (so called "salting out" salts) due to preferential hydration of the salt ions. In contrast, certain anions such as thiocyanate and perchloride and cations such as Li^{+} can increase the solubility as well as reduce the stability of a wide variety of macromolecules.

All of the above phenomena can also be considered in the light of the "distribution of microstates" picture. Thus, particular microstates or populations of microstates can be considered to produce distinct surface properties, solubilities or aggregative tendencies that are responsible for their physical degradation. Whichever description is used, however, the challenge for formulation science is to reduce the rate (and extent) at which they appear.

Nucleic acids also undergo a variety of physical changes that can be damaging to their use as vaccines, gene therapy agents, and RNA-based therapeutics.[58,98] Physical pathways of degradation of DNA and RNA are at least partially a function of their size. DNA is often used in the form of supercoiled (sc) plasmids. If the ends of a large piece of double-stranded DNA are covalently joined, it is possible for one strand to pass through the other. Assuming no breakage of either strand, the number of times this occurs is a constant known as the linking number (Lk). One strand can, however, be twisted about the other (Tw) or a writhing about the duplex axis (Wr) can occur. The relationship between these three quantities is simple:

$$\mathrm{Lk} = \mathrm{Tw} + \mathrm{Wr}$$

Changes in these parameters can occur as a result of strand breakage and reclosure (often due to enzymatic processes by topoisomerases or by physical processes as well), producing a variety of topological forms that can be easily seen as individual bands on agarose gels (Fig. 21-8) or peaks in HPLC analysis. If cleavage of a single strand occurs, a closed circular form of the DNA (oc) will be produced. If cleavage of both strands occurs near one another, linear strands of ds DNA (l) result. These forms are also easily detected by various electrophoretic methods. Because sc, oc, and l forms of DNA may well have distinct (and possibly altered) stabilities and biological activities, they represent important degradation products.

If heat is applied to either single or double stranded oligonucleotides, changes in state will also occur. In the case of large DNA molecules like plasmids, thermal melting (i.e., complete disruption of base hydrogen bonding and stacking interactions) appears only at higher temperatures (>90°C) and is probably of limited pharmaceutical relevance. An exception is longer single-stranded nucleic acids which may contain extensive regions of internal base pairing which lead to stem-loop type–structures. These regions can often melt in a highly cooperative manner with simple UV absorbance as well as other types of spectroscopic and calorimetric measurements able to detect their presence as lower temperature melting events. Shorter DNA and RNA oligonucleotides, of course, melt at much lower temperatures and this can also constitute important physical degradation events. Furthermore, "breathing phenomena" (see above) before melting and destacking of bases in single-stranded (unbase-paired) nucleic acids also comprise physical events of potential relevance to stability. It is now recognized that many RNA molecules also contain biologically significant tertiary structure due to distant contacts within polynucleotide chains (often mediated by metal ions). Although this has yet to constitute a major pharmaceutical issue, it can be expected to become so in the near future.

Nucleic acids are also capable of aggregation when the charges on the phosphate backbone of polynucleotides are partially or fully neutralized by a variety of cationic molecules such as polyamines and metals. Under such conditions, the nucleic acids may begin to associate and ultimately extensively

aggregate. This may also be a result of structural changes in the nucleic acid. Related phenomena are often

seen when DNA is complexed to polycations to create gene and vaccine delivery complexes when the zeta potential of the complex approaches neutrality. The conformational properties and stability of the DNA within gene delivery vehicles are also subject to conformational alterations, which can be accessed by methods such as CD, FTIR, and DSC. One form of such change is known as condensation. When there is a reduction in the repulsive electrostatic forces between phosphate groups, large DNA molecules can often collapse into highly condensed structures, which are of both dramatically minimized volume and high density. Such structures can possess unique physical and spectral properties. The former makes these candidates for various types of delivery applications and the latter fairly easily recognized by a variety of experimental methods (especially CD).

Chemical Degradation[99,100]

Protein drugs are often also chemically unstable with chemical degradation events occurring at specific amino acids within proteins. These degradation reactions are influenced by intrinsic factors such as primary, secondary, tertiary, and quaternary structures as well as extrinsic factors such as pH, temperature, buffer, and excipients. Common chemical degradation reactions include deamidation, hydrolysis, oxidation, *N,O*-acyl migration, and beta-elimination. This section describes the characteristics of these degradation reactions.

Deamidation Reactions

The deamidation reaction is one of the most studied and best understood reactions in peptides and proteins (Fig. 21-23). It occurs primarily in asparagine (Asn) residues. In this reaction, asparagine residues can be converted to succinimide (Asu), aspartic acid (Asp), and iso-aspartic (Iso-Asp) acid[101,102,103,104] moieties. The rate and mechanism of deamidation reactions are strongly influenced by the pH of the solution.[102,103,104] At pH < 4.0, the amide group on the side chain of the Asn residue undergoes direct hydrolysis to release ammonia to generate Asp. At pH > 6.0, the deamidation reaction proceeds via a cyclic imide intermediate (Asu) due to the attack on the carbonyl carbon of the Asn side chain by the backbone amide nitrogen of the C-terminal of the Asn residue to expel an ammonia molecule. The cyclic imide can be hydrolyzed at two different sites to produce Asp or Iso-Asp residues in proteins (Fig. 21-26). The Asp-containing protein can be further hydrolyzed at the backbone to give two protein fragments (Asp-mediated degradation). The hydrolysis of cyclic imide intermediates generates the Iso-Asp- and the Asp-containing proteins with a ratio around 4:1, suggesting that the hydrolysis reaction favors the backbone carbonyl over the side chain carbonyl group. The bulkiness of the side chain of the amino acid C-terminal to the Asn residue (i.e., the $n + 1$ residue) affects the rate of deamidation; the presence of a bulky hydrophobic side chain at the $n + 1$ residue impedes the deamidation reaction and changes the Iso-Asp:Asp ratio from 4:1 to 2.5:1. The conversion of Asn to the Iso-Asp or Asp residue adds an additional negative charge to the protein, which can potentially influence the structural and physical stability of the protein and its biological activity. This change can be detected by both peptide mapping and isoelectric focusing.

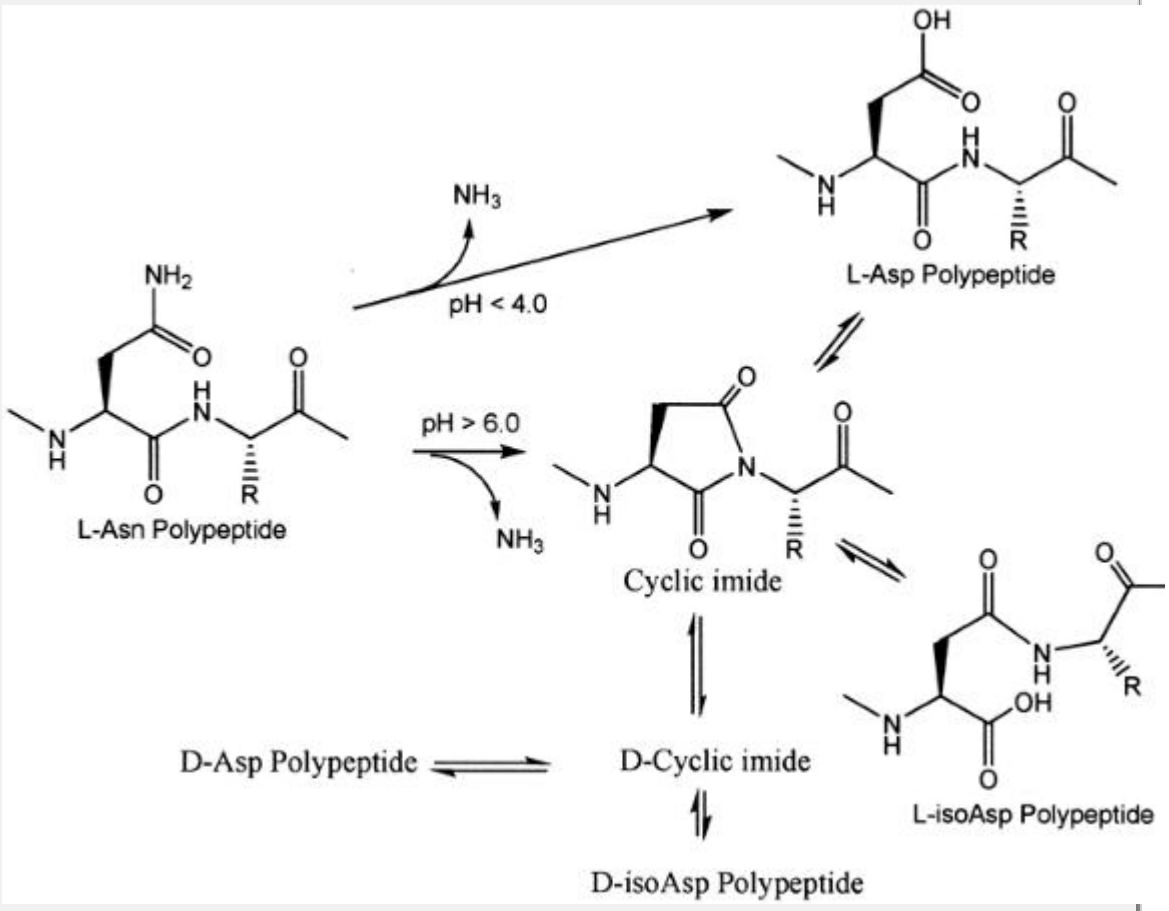

Fig. 21-23. Diagram of a deamidation reaction at Asn residues. The formation of a cyclic imide intermediate is shown followed by formation of Asp and Iso-Asp.

The impact of secondary structure on the deamidation reaction has been elucidated in both peptides and proteins; for example, bovine growth releasing factor (bGRF) peptide (Leu^{27}bGRF) contains 32 amino acids with two segments of α-helix at Phe6-Gly15 and Arg20-Leu27 (Fig. 21-24).105 The Asn8 residue on Leu^{27}bGRF undergoes a deamidation reaction to produce isoAsp and Asp peptides. To test the effect of secondary structure on deamidation, the Gly15 residue was mutated to Ala15 and Pro15 to give the $Ala^{15}Leu^{27}$bGRF and $Pro^{15}Leu^{27}$bGRF peptides, respectively. Because alanine is a strong α-helix inducer, the $Ala^{15}Leu^{27}$bGRF peptide

possesses helical structure from Asn8 to Gln30, which is a greater helical content than in the parent Leu^{27}bGRF.105,106 In contrast, $Pro^{15}Leu^{27}$bGRF does not have any α-helical structure (i.e., it is random) because proline is a strong α-helix breaker.105 Evaluation of the deamidation reaction at the Asn8 residue showed that the rate of deamidation of $Ala^{15}Leu^{27}$bGRF was slower ($t_{1/2}$ = 21.53 ±2.83 hr) than the parent peptide ($t_{1/2}$ = 15.74 ±2.45 hr); this is presumably due to the greater α-helical structure in $Ala^{15}Leu^{27}$bGRF than in Leu^{27}bGRF. Conversely, the rate of deamidation of $Pro^{15}Leu^{27}$bGRF ($t_{1/2}$ = 10.78 ±2.95 hr) is faster than the parent peptide,105,106 indicating that the secondary structure of this peptide has an accelerating effect on the deamidation rate.

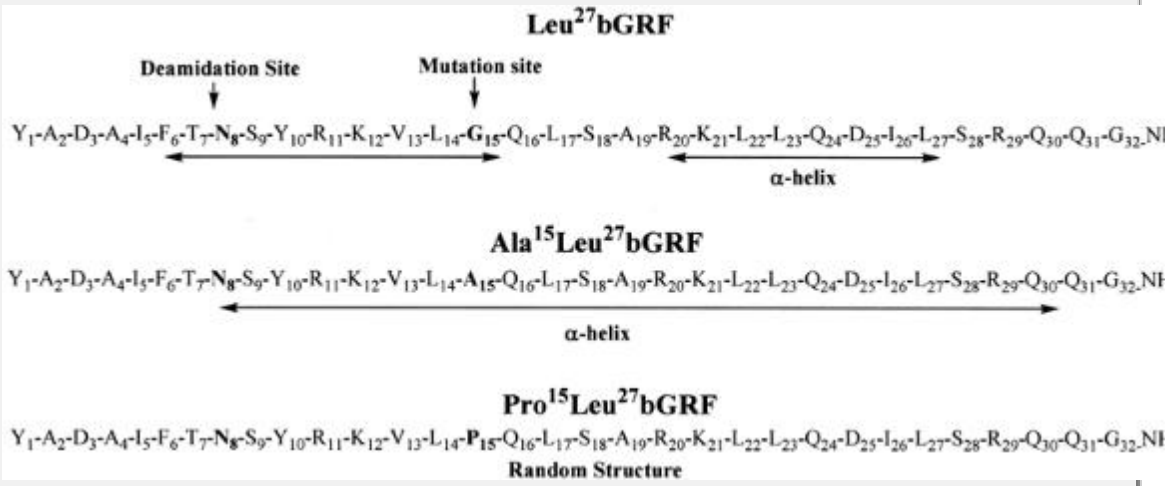

Fig. 21-24. The sequence of Leu^{27}bGRF and its mutants Ala^{15}Lue^{27}bGRF and Pro^{15}Leu^{27}bGRF. Asn8 is the deamidation site on these peptides. The mutations discussed in the text affect the secondary structure of the peptides.

The rate of deamidation reactions is also affected by the position of Asn residues in β-turns. Linear (AcNG and AcGN) and cyclic (cNG and cGN) peptides have a β-turn structure at the Asn-Gly or Gly-Asn sequence in which the Asn residue is at the *n* + 1 or *n* + 2 position of the β-turn (Fig. 21-25).107 At pH 8.8, the rates of degradation of the cyclic peptides cGN (<2.2 × $10^4 s^{-1}$) and cNG (9.36 × $10^7 s^{-1}$) are slower than the respective linear peptide counterparts, AcGN (20.1 × $10^7 s^{-1}$) and Ac-NG (42.2 × $10^7 s^{-1}$) [107]. This result suggests that the molecular rigidity of the cyclic peptides hampers the deamidation reaction. In addition, the cGN peptide with Asn at *n* + 2 is more stable than the cNG peptide with Asn at *n* + 1. The Asn residue at *n* + 1 of a β-turn undergoes deamidation reaction more readily than the Asn residue at the *n* + 2 position; this is due to a more favorable formation of the cyclic imide when the Asn residue is at the *n* + 1 position than when it is at the *n* + 2 residue of a β-turn. For the Asn residue at the *n* + 1 position, the shortest distance between the carbonyl carbon of the Asn side chain and the backbone nitrogen atom of the *n* + 2 residue for forming the cyclic imide intermediate is 1.89 Å (Fig. 21-23). In contrast, the shortest distance between reactive atoms when the Asn residue is at position *n* + 2 is 4.8 Å (Fig. 21-23). In conclusion, the secondary structure and the position of the Asn residue in a peptide or protein affect the rate of their deamidation reactions. Furthermore, it has been demonstrated that increased mobility of deamidation sites in proteins (e.g., when they are located on a protein's surface) also facilitates the deamidation reaction.108

Asp-Mediated Backbone Hydrolysis

Aspartic acid residues can catalyze backbone hydrolysis in peptides and proteins.109,110,111,112,113 The deamidation products of a protein containing susceptible Asn residues are Asp- and

iso-Asp–containing proteins that can undergo peptide bond hydrolysis at both the N- and C-terminal of the Asp residue via pathways a and b, respectively (Fig. 21-26). In pathway a, the carbonyl carbon of the *n* - 1 residue is attacked by the carboxyl group of Asp to form a six-membered ring intermediate.112 Upon rearrangement, the six-membered ring is open to the anhydrate intermediate that immediately hydrolyzes to form two fragments of the peptide or protein. Pathway b proceeds via the attack of the C-terminal of the carbonyl of the Asp residue by the side chain carboxylic acid oxygen to form a five-membered ring. The rearrangement of the five-membered ring hydrolyzes the peptide bond

to separate the two portions of the protein. Like the Asn residue, Asp can also form a cyclic imide intermediate that can further rearrange to form iso-Asp. Comparison of the stability of Asp-containing linear and cyclic peptides has shown that the rate of peptide bond hydrolysis mediated by the Asp residue in a cyclic peptide is slower than the rate of degradation in a linear one.112Molecular dynamics simulations show that it is more difficult to form cyclic imide intermediates in a cyclic peptide than in a linear peptide; this is due to the rigidity of the cyclic peptide backbone and a favorable distance between the reactive atoms which can be easily accommodated in the linear peptide compared to that of cyclic peptides.113

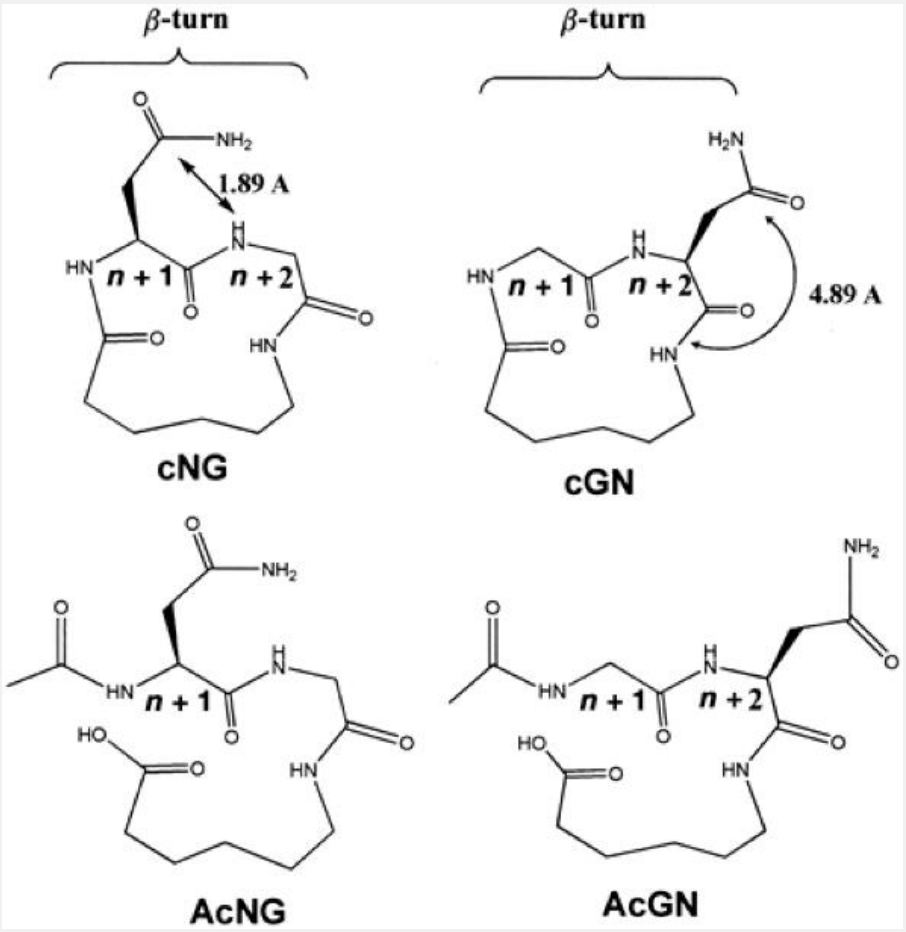

Fig. 21-25. The effect of the position of the Asn residue in a β-turn on a deamidation reaction.

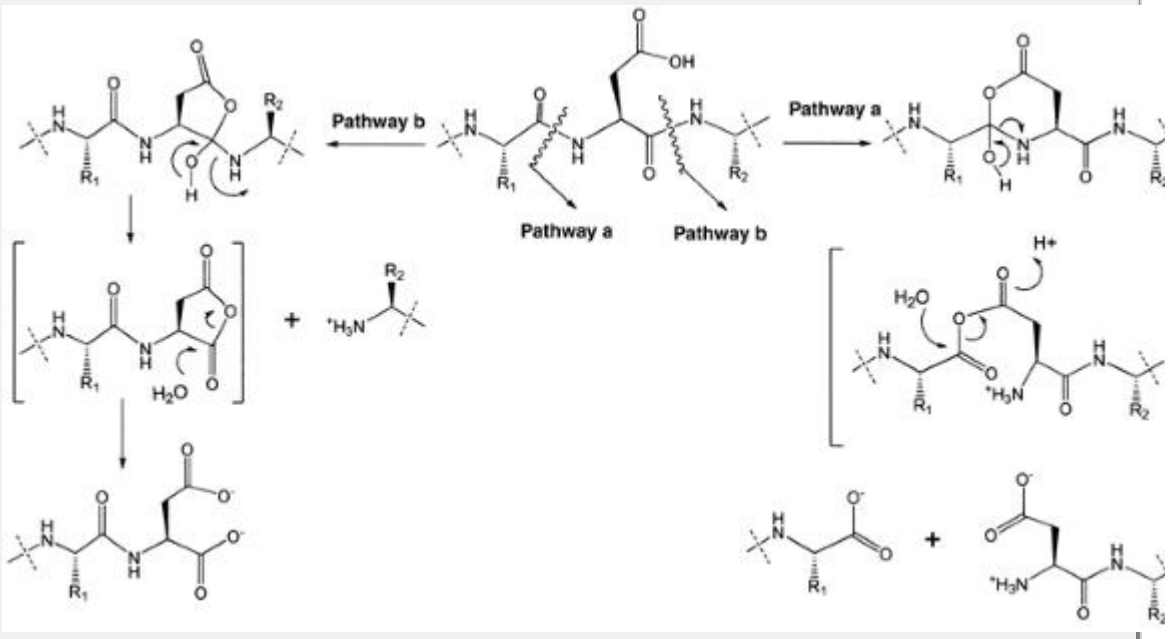

Fig. 21-26. The degradation pathways of Asp residues to induce peptide bond hydrolysis.

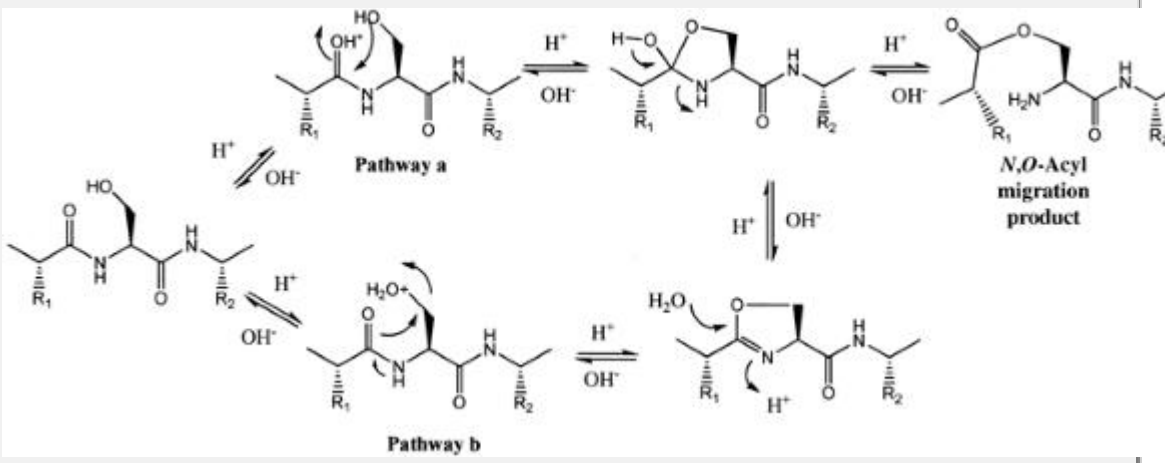

Fig. 21-27. The *N,O*-acyl migration reaction occurs at Ser and Thr residues. This reaction produces a rearrangement of the peptide backbone at the N-terminus of the side chain of the Ser or Thr residue to make an ester bond.

N,O-Acyl Migration in Ser or Thr Residues

Ser and Thr residues are prone to undergo *N,O*-acyl migration reactions which rearrange the protein/peptide backbone in acidic conditions (Fig. 21-27).114 This reaction can occur

via two possible mechanisms (pathways a and b). Pathway **a** is initiated by protonation of the carbonyl oxygen of the residue *n* - 1 to the reactive Ser residue followed by an attack of its carbon by the Ser

residue's oxygen to form a five-membered ring intermediate. The opening of the five-membered ring upon cleavage of the C–N bond generates an ester bond from the carbonyl of the (*n* - 1) residue to the side chain oxygen of Ser. The second possible mechanism is via protonation of the hydroxyl group of the Ser residue followed by an attack of the beta-carbon of the Ser residue by the carbonyl oxygen of the *n* - 1 residue to produce a five-membered oxazoline, which upon the nucleophilic attack of water on the double bond produces the five-membered ring oxazolidine intermediate. As in pathway a, opening of the five-membered ring intermediate generates the *N,O*-Acyl migration product.

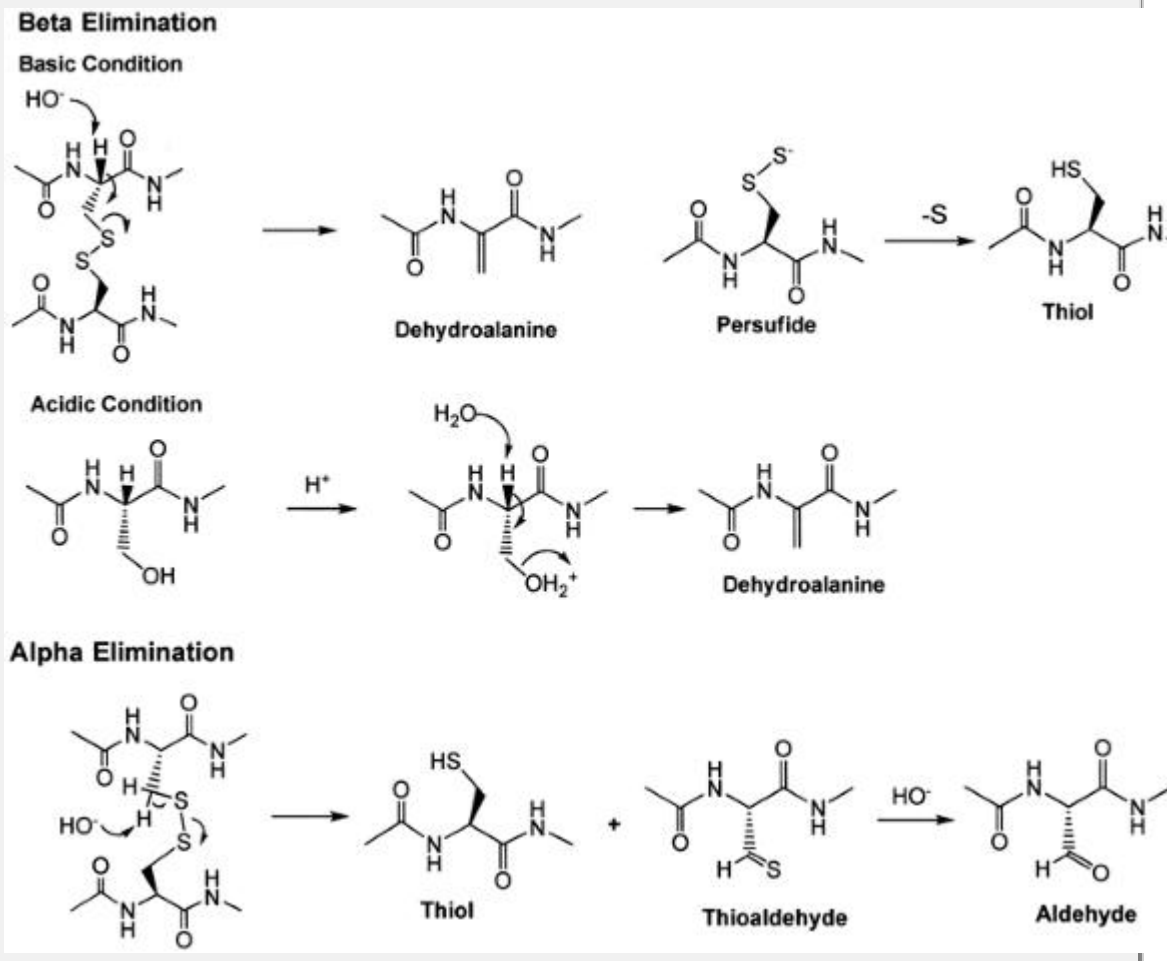

Fig. 21-28. The beta- and alpha-elimination reactions in Ser and Cys residues under acidic and basic conditions.

Beta- and Alpha-Elimination Reactions

Disulfide bonds have important roles in stabilizing the folded structure of proteins. They are normally formed by two Cys residues that are in close proximity due to tertiary structure constraints. The destruction or reduction of disulfide bonds may frequently have an impact on the structure and biological activity of a protein. The degradation of disulfide bonds can occur in mild to strong alkaline conditions when hydroxide ions abstract the alpha-proton of the Cys residue to generate dehydroalanine and persulphide ion (Fig. 21-28).[112,115] Extrusion of the sulfur atom from the persulfide ion produces the thiol group of a Cys residue. In basic conditions, beta-elimination is often observed in Cys residues that are involved in disulfide bonds. This reaction is frequently observed in proteins that contain disulfide bonds. In acidic conditions, beta-elimination can take place in Ser residues; upon protonation of the side chain OH group, the alpha proton of Ser is abstracted by a water molecule to produce the dehydroalanine residue (Fig. 21-28).

Alpha elimination can also occur in a disulfide bond under basic conditions to form thioaldehyde and aldehyde products. The alpha-elimination reaction proceeds via proton abstraction of the beta-carbon of the Cys residue to form thio-aldehyde and releases the thiolate anion from the other Cys residue. The thio-aldehyde can further react with the hydroxide anion to produce aldehyde. The presence of a reactive aldehyde may further react with amino groups (i.e., Lys side chains) within a protein or with another protein to form an imine bond.
P.550

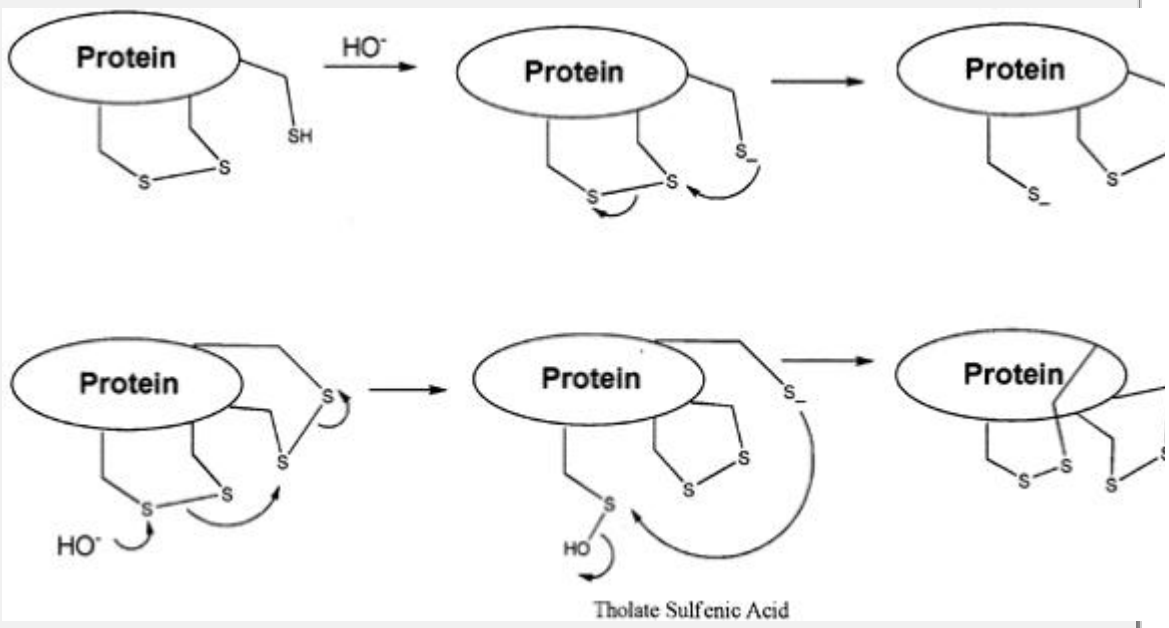

Fig. 21-29. Disulfide bond exchange in proteins under basic conditions.

The presence of multiple disulfide bonds in a protein can also lead to a disulfide bond exchange reaction under basic conditions.116 At high protein concentrations, disulfide bond exchange can form dimers and higher oligomers that can precipitate the protein from solution. The exchange reaction can be initiated by the attack of thiolate anion on the sulfur atom of a disulfide bond (Fig. 21-29). Presumably, this exchange reaction occurs when the thiolate anion is in close proximity to the disulfide bond. The exchange reaction could also occur via the attack of the hydroxyl anion of the sulfur atom of the disulfide bond to produce thiolate anion and sulfenic acid. Further reaction of the thiolate anion with the sulfur of sulfenic acid to release hydroxyl anion can produce the disulfide bond exchange reaction.

Oxidative Reactions (Met, His, Trp)

Oxidation reactions of methionine (Met), histidine (His), and tryptophan (Trp) residues are often observed during protein production and formulation. Such oxidation reactions are due to reactive oxygen species (e.g., $^{\bullet}OH$, $O_2^{\bullet-}$, H_2O_2, O_3, $^1O_2$). The formation of reactive oxygen species can be catalyzed by metals (i.e., ferryl, perferryl) and can be produced by ionizing radiation and photochemical reactions. To prevent protein drug oxidation, reducing agents (i.e., glutathione, dithiothreitol, thioacetic acid, and cysteine) have been added to protein formulations. Methionine oxidation produces the sulfoxide amino acid and further oxidation generates a sulfone group on the side chain of Met (Fig. 21-30a).[117,118,119,120] The sensitivity of the Met residue to oxidation depends on its location within the tertiary structure of a protein. In protein formulations, the oxidation reaction can occur because of the presence of residual hydrogen peroxide that is used to sterilize containers and vials for storage. Oxidation of methionine can significantly reduce the half-life of protein drugs and generate major

problems in protein purification and formulation. Oxidation may also alter the physical stability of proteins through the production of oxygen radicals.120,121

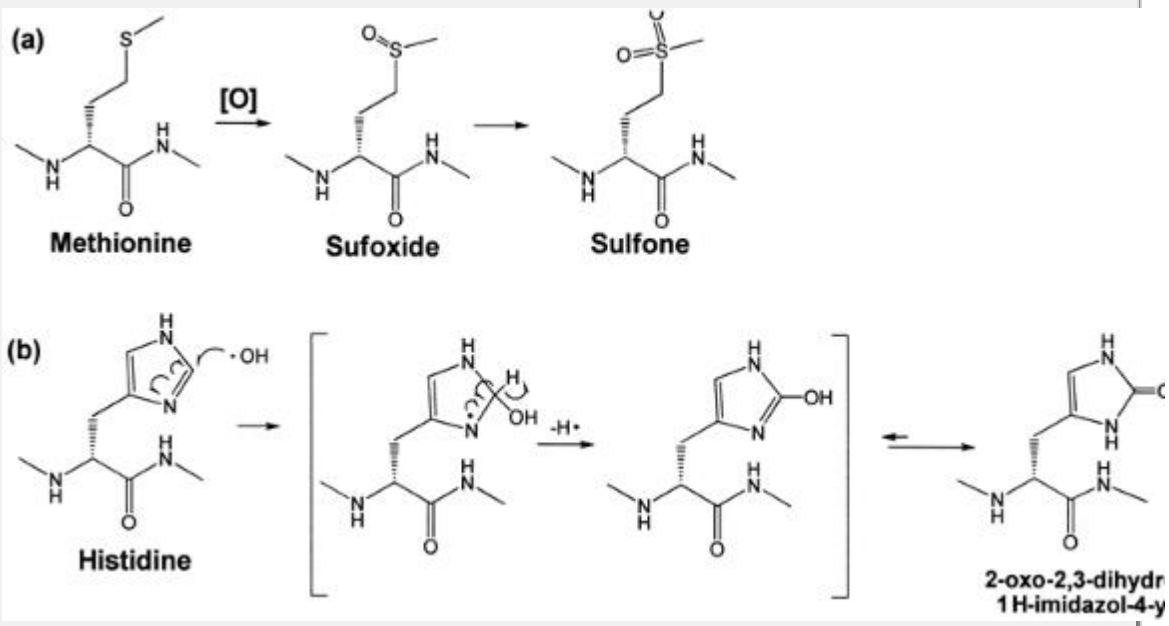

Fig. 21-30. (*a*) The oxidation of methionine to form sulfoxide and sulfone. (*b*) Oxidation of a histidine residue by hydroxyl radicals to form oxo-imidazole.

The aromatic rings of His, Trp, Tyr, and Phe residues are also prone to oxidation. The imidazole ring of His in serum albumin is oxidized by ascorbic acid/Cu^{2+} or H_2O_2/Cu^{2+}, producing an oxo-dihydro-imidazol ring (Fig. 21-30b).122 This oxidation is via a hydroxyl radical attack at the C2 position of the imidazole ring followed by the removal of the hydrogen radical to produce the 2-oxo-imidazol moiety. Further oxidation of His can produce an Asp residue.122

The oxidation of Trp residues in peptides and protein with hydrogen peroxide generates *N*-formylkynurenine (NFK), kynurenine (Kyn), oxindolylalanine (Oia), dioxindolylalanine (DiOia), and 5-hydroxytryptophan (5-OH-Trp).123,124,125 Similarly, oxidation of the Trp residue in di- and tripeptides (i.e., Ile-Trp, Trp-Leu, Gly-Trp-Leu, and Ala-Trp-Ile)

by a superoxide-generating system such as hypoxanthine/ xanthine oxidase in the presence of Iron(III) and ethylenediaminetetraacetic acid (EDTA) generates NFK and Oia as the major products. The hypoxanthine/xanthine oxidase/Fe(III)-EDTA system generates the reactive oxygen species hydroxyl radical ($^{\bullet}OH$) produced via a Fenton reaction that oxidizes the Trp residue.124

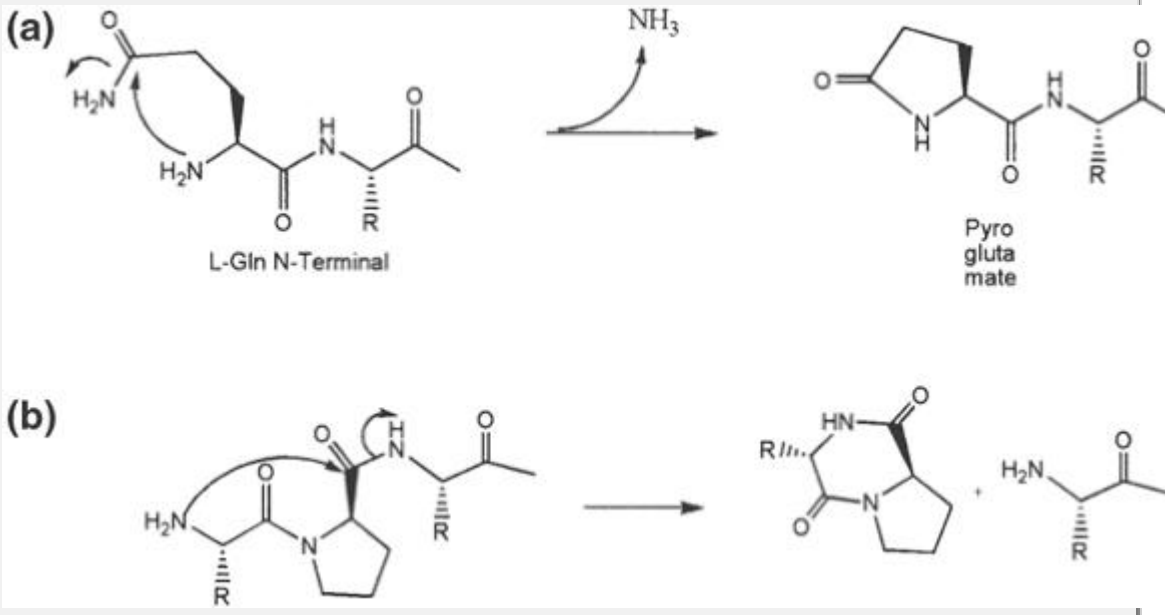

Fig. 21-31. (*a*) The conversion reaction of Nterminal glutamine to pyroglutamine. (*b*) The reaction to form diketopeperazine at the N-terminal of a protein or peptide.

Reactive oxygen species can oxidize the aromatic ring of Tyr to produce 3,4-OH-Phe as the major product; this reaction can also occur upon exposure of proteins to ionizing radiation. Tyrosine-tyrosine cross-linking has been observed in proteins upon UV and gamma irradiation as well as in low-density lipoproteins found in vivo.126 Similarly, the aromatic ring of Phe can be oxidized to form Tyr, 2-OH-Phe, 3-OH-Phe, and 2,3-OH-Phe with *ortho* tyrosine (2-OH-Phe) as the major product.

Other Reactions

Other side reactions in peptides and proteins include the formation of pyroglutamate and diketopiperazine. Peptides and proteins that have glutamine and glutamic acid residues at their N-terminus can form pyroglutamate degradation products (Fig. 21-31). The deamidation of glutamine at the N-terminus is more rapid than the deamidation of this residue when it is located in the middle of protein sequences. The driving force for this reaction is the formation of a stable five-membered ring when the Gln or Glu residue is present at the N-terminus. This reaction is not observed in N-terminal Asn residues because the ring product is an unfavorable four-membered ring. Pyroglutamate formation was observed during a stability study of a decapeptide vaccine (ELAGIGILTV) containing an N-terminal glutamic acid.127

Peptides and proteins that have an Xaa-Pro residue at the N-terminus may be prone to diketopiperazine formation with release of the rest of the protein product with deficient Xaa-Pro residues. This reaction was first observed in a Gly-Pro peptide (Fig. 21-31). This reaction has been observed in recombinant human vascular endothelial growth factor.

As mentioned previously, oligonucleotides (i.e., DNA and RNA) have also been investigated as potential therapeutic agents. These molecules are also subject to a variety of chemical degradation reactions.100'98 DNA and RNA molecules can both undergo various chemical changes via hydrolysis or oxidation reactions. RNA is generally less stable than DNA. The hydrolysis reaction can cause the breakup of the oligonucleotide chain and isomerization of the phophoester group on the ribose ring. As expected, these degradation reactions are strongly influenced by the intrinsic properties of the solution such as pH, buffer, and ionic strength. Similar to proteins, external conditions such as temperature and light can also have dramatic effects on the stability of oligonucleotide-based drugs. Physical instability

(i.e., conformational changes, aggregation, and precipitation) of oligonucleotides and plasmids can often be induced by their chemical degradation.

The hydrolysis reactions of oligonucleotides are catalyzed by acid or base as illustrated for RNA degradation in Figure 21-32. Acid-catalyzed degradation produces two degradation pathways. The first results in a phosphoester bond shift from C3′ to C2′ to make oligonucleotide 5; this bond shift will affect the higher-order structure of the RNA. In this case, the acid-catalyzed degradation is initiated by protonation of the oxygen of the phophoester group (compound 2) followed by a nucleophilic attack of the 2′-OH group on the phosphorous atom to produce intermediate 3. The proton transfer from the 2′-oxygen to the 3′-oxygen generates intermediate 4, which upon a five-membered ring opening reaction produces degradation product 5 with an oligonucleotide chain shift. The second pathway generates two fragments of RNA (i.e., 7 and 8 in Fig. 21-32). In this route, the proton on the 2′-oxygen in compound 3 can shift to the oxygen attached to the methylene group of the next nucleic acid to generate intermediate 6. The unstable intermediate 6 undergoes a fragmentation reaction and produces the two smaller pieces of RNAs, 7 and 8.

Fragmentation of RNA can also be catalyzed by base. In this case, the reaction is initiated by deprotonation of the 2′-hydroxyl group as shown in intermediate 9 followed by the nucleophilic attack of the phosphorous atom by the 2′-oxy-anion to yield intermediate 10. The base abstraction of the hydroxyl proton of the phosphate group in intermediate

10 leads to the fragmentation of the RNA to give a smaller RNA 11 and intermediate 12. Opening of the five-membered phophoester in 12 produces another RNA fragment 13.

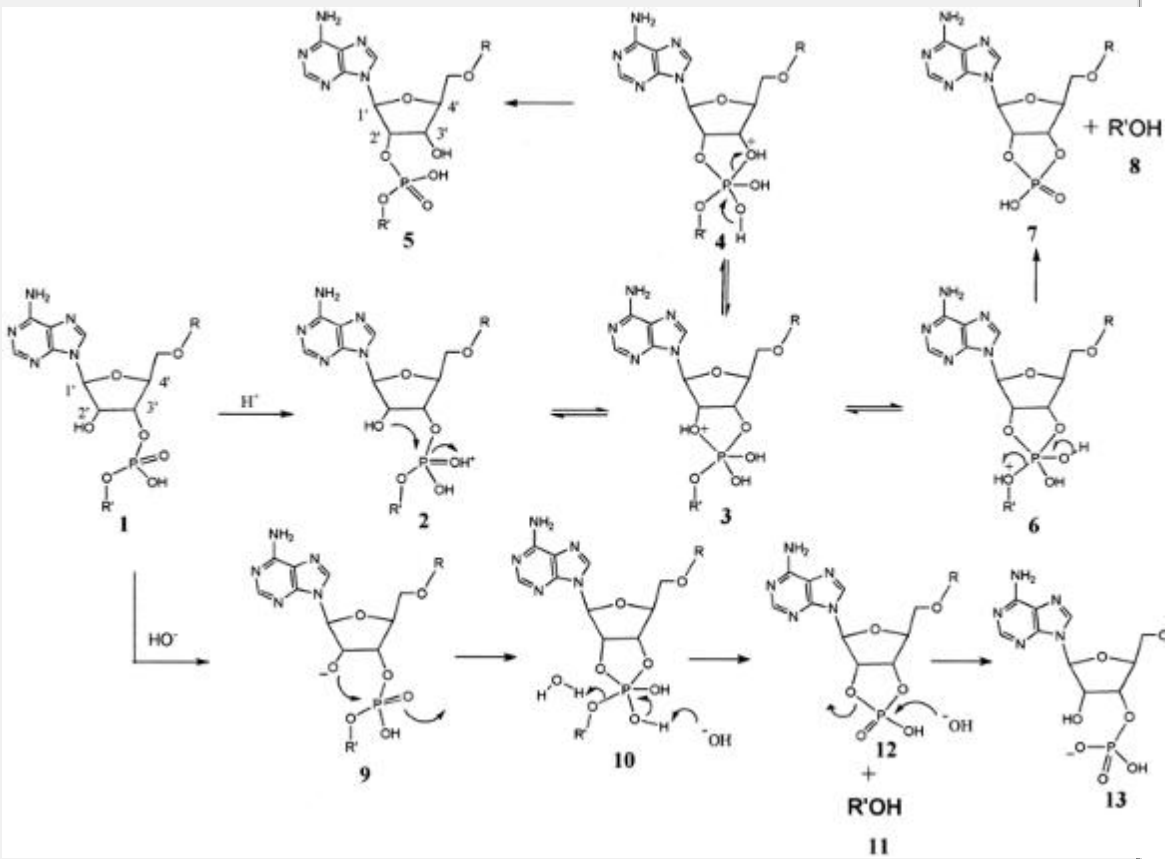

Fig. 21-32. Schematic diagram of some degradation reactions of RNA catalyzed by acid or base. These degradation reactions cause RNA fragmentations and

phosphoester bond shift.

Oligonucleotides undergo other hydrolysis reactions at different locations within the molecule to produce other products, including the release of a base (e.g., purines) as well as the ring opening of the ribose group. Furthermore, modification of the base groups can be catalyzed by base. Oxidation reactions of RNA produce RNA with an open ribose ring as well as RNA with modified base groups. For a further description of these other reactions and related changes in DNA, readers are encouraged to read the appropriate reviews.[100,98]

Formulation

In the previous sections the production, characterization, and the most common physical and chemical pathways of degradation of biotechnology-based products were described. How are these macromolecules and their complexes formulated? The goal is to take the various types of information that was gathered to create actual drugs and vaccines.

The initial major concerns with solution state biopharmaceuticals usually involve conformational stabilization, prevention of aggregation, and inhibition of chemical degradation reactions. Thus, the first step in most formulation procedures is to identify such events. This is most commonly done using "accelerated-stability" protocols. Thus, various types of stress are applied to the system of interest and the methods described previously are used to detect changes in their physical and chemical structure including the association state of the molecule/macromolecular complex. The most common forms of stress applied in rough order of their utility are temperature, pH, redox potential, solute, shear (shaking), and freeze/thaw cycles. Ideally, one could employ actual, selected storage conditions (e.g., 2°C–8°C, 24–48 months) but this is generally not possible due to temporal constraints in product development timelines. Ultimately, however, real-time stability studies must be the ultimate arbitrator of successful stabilization. Given the molecular complexity of biopharmaceuticals and the wide variety of methods available for their analysis, the decision on how to proceed is often a difficult one.

One frequent approach is to pick one or several techniques that are expected to be sensitive to major degradation pathways. For example, one might use DSC to evaluate thermal stability at several different pH values and monitor stress-induced aggregation with SEC, oxidation induced by H_2O_2, and deamidation by high pH with HPLC-MS. Then, as described below, potential excipients can be tested for their ability to inhibit any degradation processes observed. The problem with such an approach is that important degradation

events might be missed because of the limited ability of a small number of conditions, events, and methods to detect all potential problems. There are a large number of variations on this approach based on both the techniques and types of stress employed. Alternatively, attempts can be made to cover a much wider formulation space by using many techniques and a wide range of solution conditions. One widely documented approach to the analysis of physical degradation makes measurements at 0.5 or 1.0 pH intervals from 3 to 8 and from low to high temperatures (e.g., 0°C–100°C). A series of methods such as CD (to detect secondary structure changes), intrinsic fluorescence (to monitor tertiary structure), ANS fluorescence (to detect alterations in apolar surface exposure), and static and/or dynamic light scattering to measure association (aggregation)/dissociation phenomena are used to characterize the response of the physical state of the system to stress. Changes in dynamic behavior can be analyzed in a similar manner using methods such as isotope exchange, US spectroscopy, PP-DSC, fluorescence anisotropy, or solute quenching. Similarly, the presence of chemical alterations of individual residues can be described in terms of their rate constants as detected by LC-MS. All of the above approaches have recently been facilitated by the availability of HTP technology for their execution. The major weakness in all of the above is that it is still possible to miss key degradation events. Furthermore, the results obtained may not always be extrapolated to actual pharmaceutical storage conditions although this does not often appear to be the case. Variations of this HTP approach have been successfully applied to

therapeutic peptides, proteins, VLPs, viruses, and bacterial cells as well as various vaccines types and accompanying adjuvants. It can also be used with solid-state formulations (see below). How does one analyze the vast array of information that is obtained by such an extensive collection of data and experimental conditions? One method is to use the EPD method described earlier for the analysis of high-resolution UV absorption data. In the multiple techniques version, however, normalized data from all of the techniques employed are used to construct the EPD. The resultant colored summary of the effect of the chosen variables (T, pH, drug concentration, ionic strength, agitation, freeze/thaw, etc.) on the physical and chemical structure and behavior of the biopharmaceutical can then be used to guide formulation development (Fig. 21-33). We should also mention that it is possible to redesign the macromolecular system if the exact mechanism of degradation is known. For example, residues in a protein that undergo deamidation or oxidation could be replaced with nondegrading analogues or the interior packing of the protein's amino acid residues could be improved through modern protein design methods to improve stability. Although in many ways an ideal solution (although immunogenicity can become a problem), this has yet to become a routine approach to stability problems. It may become so in the future.

Often the first thing one does with a biopharmaceutical solution formulation is to select a buffer. A number of considerations are necessary for an optimal choice. First and foremost is that the buffer supports optimum structural stability and solubility. This can often be achieved through the methods outlined above. For example, a pH can be selected to place the formulation as far as possible from structural changes produced by changes in pH and temperature using a single method (e.g., CD, fluorescence, DSC) or as distant as possible from the apparent phase boundaries of an EPD. One caveat, however, is that a number of proteins undergo pH-dependent conformational changes as a part of their normal biological functions. For example, the lowering of the pH inside endosomes (sometimes the immediate destination of a protein taken up into a cell) might induce a structure change that exposes apolar regions, which permits interaction with the endosomal membrane and subsequent release into the cytoplasm. Furthermore, because so many degradation processes are pH dependent, it may be necessary to compromise in the selection of pH. Tables of physiologically acceptable buffers and their pK_a's are readily available. Among the many popular buffers used for biopharmaceuticals are phosphate (pK_a's 2.12, 7.21, 12.32), citrate (3.06, 4.74, 5.40), and imidazole (7.00). By combining buffers, it is also possible to obtain a very broad buffering range.

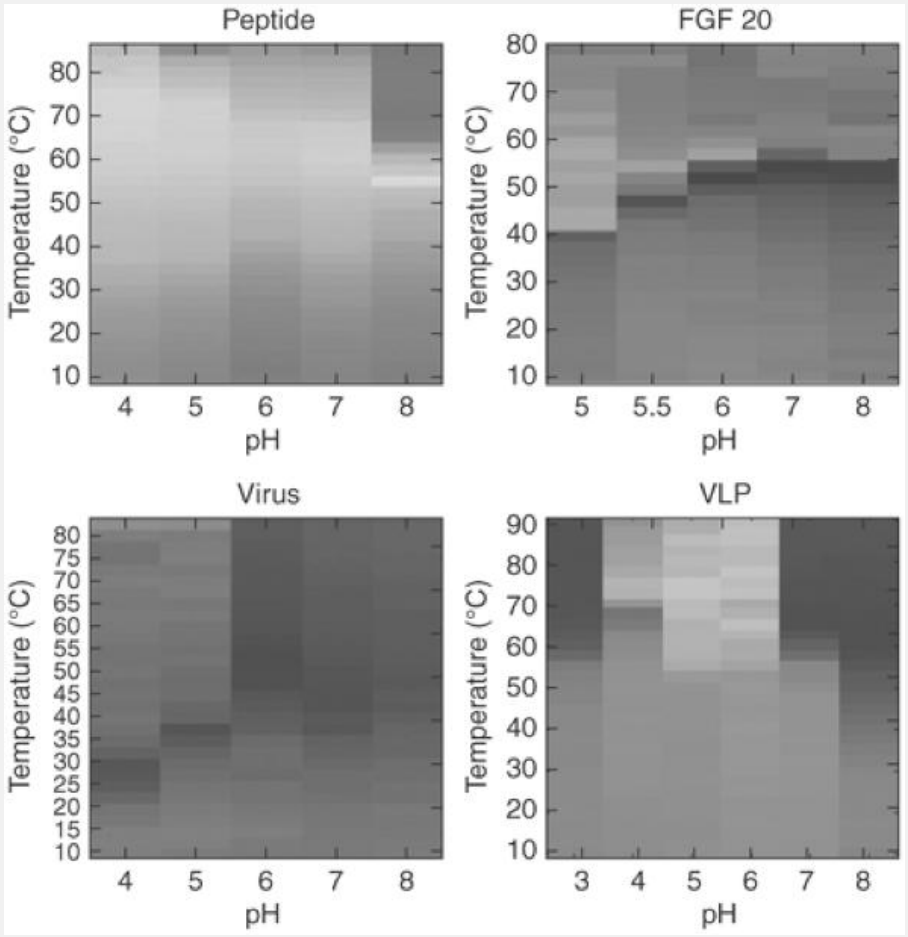

Fig. 21-33. Representative EPDs using multiple methods. For example, CD spectra (secondary structure), intrinsic fluorescence (tertiary structure), ANS fluorescence (appearance of apolar binding sites), and light scattering (association and aggregation) are obtained after normalization as vectors and represented as colors (or in this case regions of different shadings). Apparent phases of different colors then represent different structure states. See Figure 21-11 and text.

A variety of other factors may be involved in buffer choice. Buffer ions may specifically interact with proteins due to their charged nature. They may also chelate or be contaminated

with metals, a potentially important phenomenon. One may also need to consider the effect of temperature on a buffer's pK_a values. Some buffers, such as the commonly used Tris (tris-(hydroxymethyl) aminomethane) species shift their pK_a by as much as -0.03/°C.

The next step in the formulation of macromolecules (if necessary, as it frequently is) is the selection of excipients to control critical degradation processes. This is usually done by screening a group of compounds and polymers usually referred to as GRAS (Generally Regarded as Safe) materials. This can be accomplished by using any one or a combination of the methods discussed above. Potential excipients are usually initially tested at high concentrations with concentration dependence studies employed later to define the minimum concentration that can be used to obtain the desired effect.

The GRAS excipient list is based on compounds currently used in marketed formulations of drugs. It consists of a collection of carbohydrates, polysaccharides, amino acids, small molecules, detergents,

and polymers among other agents. A number of different mechanisms mediated by these compounds can stabilize macromolecular systems. High concentrations (i.e., >0.3 M) of sugars, amino acids, and some salts appear to stabilize through a mechanism known as preferential exclusion.128 This is based on the greater surface area of a macromolecule in its structurally disrupted (unfolded) state. The presence of the stabilizing agent causes the chemical potential (free energy) of the macromolecule to be increased in a manner proportional to its surface area. Because this is an unfavorable process, the effect is to differentially stabilize the native state (Fig. 21-34). In contrast, some stabilizers bind better directly to the native state. This shifts the N↔U equilibrium to the native form resulting in stabilization. For example, the presence of extended polyanion binding sites on many proteins such as growth factors and coagulation factors means that polymers like heparin and dextran sulfate can often have dramatic stabilizing abilities.129,130,131 Some compounds act by either directly or indirectly inhibiting aggregation. Those that act indirectly generally do so by stabilizing the native state which delays formation of aggregation competent species such as MG forms. Direct effects occur through blocking of the protein/protein interactions that are responsible for association processes. Inhibition of protein aggregation by detergents is thought to occur through one or both mechanisms. Because of the presence of disulfide bonds and free thiol groups in proteins, it is sometimes possible to stabilize proteins by the inclusion of a reducing agent to either maintain free thiols in their reduced (and active) form or prevent the formulation of nonnative inter- or intramolecular disulfides which leads to inactive forms. Since metals can inactivate macromolecules through a variety of mechanisms such as oxidation, the presence of a chelating agent can be used to minimize such problems.

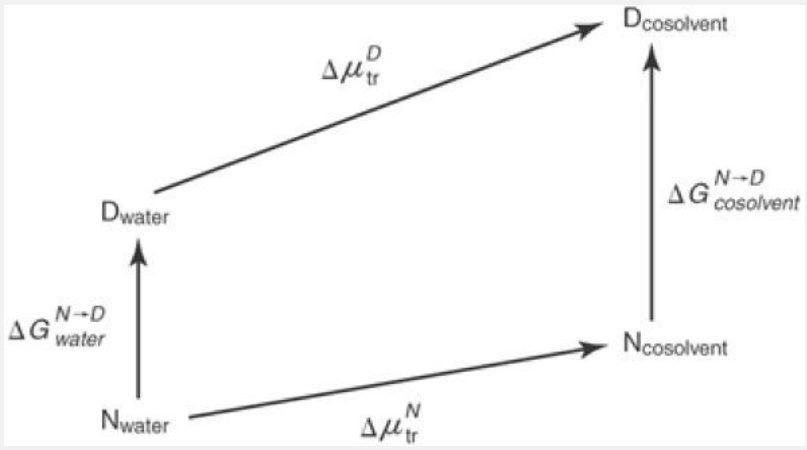

Fig. 21-34. The thermodynamic mechanism of stabilization of macromolecules by solutes which cause their preferential hydration. The solute (cosolvent) causes a greater difference in free energy of the unfolding reaction than in its absence leading to a destabilization of the unfolded (D) state (and therefore a relative stabilization of the native (N) form).

As mentioned previously, the amphipathic nature of proteins means they are usually quite surface active. By this we mean that they have a strong tendency to bind to surfaces such as air/water interfaces as produced by agitation or to the inner surfaces of storage devices such as vials and syringes. In fact, at low concentration (<10 µg/mL), a substantial portion of the macromolecules may be resident on a surface.132 At least three common approaches have been used to minimize these problems. In the first, the design or nature of the surface itself can be altered. This is accomplished most frequently by the use of different materials or the addition of a coating that lowers interactions with proteins. The second method often employs the presence of a proteinaceous material such as serum

albumin, casein, or gelatin to competitively prevent or displace the macromolecular drug substance or vaccine from the surface. Initially, animal-derived versions of these proteins were used but they are being replaced by recombinant forms of the same or similar proteins. Third, detergents are often used for a related purpose due to their affinity for both proteins and/or container surfaces. It has also recently been recognized that particles originating from various sources such as the plastic or metal materials produced by the degradation of vial filling pumps and the tungsten used in syringes can result in particulate matter that must be removed for clean formulations to result. Many of the above phenomena can also result from conditions produced during the shipping of biopharmaceuticals. Solutions to such problems are similar to those described above, but careful shipping studies are essential to identify and minimize their occurrence.

Drying of Biopharmaceuticals133

Despite recent advances in the preparation of stable, solution-based formulations of peptides, proteins, nucleic acids, and viruses, it is still frequently necessary to employ dried formulations of biopharmaceuticals. By removing most (but not all) of the water and reducing inter- and intramolecular mobility, it is usually possible to dramatically stabilize such systems. Dried formulations are often considered less desirable because of significantly increased expense and the need

for reconstitution, but the dramatic improvement in stability obtained by such technologies is thought to be more than adequate reason for their use. In fact, many currently marketed biopharmaceuticals and vaccines employ such methods. By far, the most common method employed is that of lyophilization, which is also known as freeze-drying. The techniques of spray drying and spray–freeze drying have, however, also been used to dry biopharmaceuticals. There exist a variety of other potentially useful technologies such as foam drying, which will not be described here.

Freeze-drying is far and away the most commonly used procedure.134 It is essentially a batch process in which water is removed directly from the solid state. It is generally performed in three distinct phases. In the first, the water in a solution of the biopharmaceutical agent is converted to ice. This results in the production of a concentrated frozen macromolecular solid. This freezing step is typically performed in the temperature range of -45°C to -10°C for 2 to 5 hr. The second stage is referred to as primary drying. This involves the removal of some unfrozen water (ca. 15%) and sublimation of ice at -10°C to -40°C for 5 hr to 5 days (this stage is highly variable in time). The final procedure (secondary drying) involves removal of most of the remainder of the unfrozen water down to 1% to 4% as the temperature is increased from the previous process to 4°C to 50°C for 5 to 15 hr.

In contrast, spray drying is a continuous process and involves drying from the liquid state.135 Its initial step consists of atomization of a macromolecule containing aqueous solution into small droplets. This is generally considered to be the most problematic aspect of the procedure given the presence of the air/water interface, a potential site of protein degradation and aggregation. The droplets are mixed with hot air (ca. 120°C) which rapidly (in seconds) removes most of the water resulting in concentrated solute. After further cooling at 40°C to 50°C, only a low water content (3%–5%) remains. The latter water content is typically somewhat greater than that produced by lyophilization and may result in greater degradation of the macromolecular drug or vaccine upon storage. In spray-freeze drying, the atomization and freezing is carried out in a solvent such as liquid nitrogen followed by macromolecule/vaccine drying.

Because it is so much more commonly employed, we will consider lyophilization in more detail. In the initial freezing step, most biomolecules form amorphous solids (in contrast to many small molecules solutes which may crystallize). In general, primary drying is performed 2°C to 3°C below T_g' (Tee Gee prime), the glass transition temperature of the freeze concentrate. The glass transition temperature (T_g for a pure solid) refers to the softening of a glasslike solid to form a viscous liquid state, which permits increased molecular mobility and subsequently enhanced degradation. One potential problem occurs because of the concentration that occurs during freezing. A buffer-like sodium phosphate may crystallize causing a shift of several units to lower pH, a potentially degradative condition. Thus,

phosphates are usually not employed (although small amounts may be acceptable). In general, a minimal weight ratio of buffer to other solutes is used to minimize crystallization-induced pH shifts as well as prevent large reductions in T_g' and therefore increased solute mobility.

What is the significance of glass transitions136? During primary drying (ice sublimation), T_g' reflects the temperature at which the conversion from a glassy to rubbery solid state occurs. As the material is dried it can also undergo a loss of structure that is referred to as cake "collapse" with the temperature at which this occurs designated the collapse temperature (T_c). This can be detected by a special form of microscopy known as "freeze-drying microscopy"137 or by DSC138 as a district thermal event or by a change in electrical resistance. The collapse temperature is usually a few degrees higher than T_g'. During secondary drying as the unfrozen water is removed, this phase change is seen hear the T_g. These temperatures are very formulation dependent. As one moves above a glass transition temperature, the mobility and reactivity of the macromolecule or its complexes increase. During storage, the T_g can have a controlling impact on the drug's stability. In general, as the amount of residual water increases, the T_g is lowered. Thus, knowledge of this property is one key to preparing a stable, lyophilized formulation.

To create a stable dry formulation of a biotechnology-based drug, all of the above must be considered in its creation. Besides optimization of the lyophilization cycle, such formulations almost always contain excipients.139,140 These are used to facilitate stabilization during freezing stress ("cryoprotection"), stabilization during freezing and drying ("lyoprotection"), and stabilization in the dry state to enhance integrity during storage. In addition, excipients are used for a number of other reasons. Bulking agents such as mannitol or glycine are often employed for "elegance" and to prevent "blow-out" in which the dry cake can be expelled into the freeze dryer. Bulking agents are often chosen for their crystallinity and their high eutectic temperature to facilitate rapid, easy drying. Crystallinity is typically evaluated by a combination of polarized light microscopy (to detect birefringence), x-ray powder diffraction, and calorimetry. Buffers are often included for pH control although care must be taken that their crystallization does not produce large and potentially destructive pH shifts. As a general rule, however, they are used in minimal amounts. Isotonicity modifiers such as glycerol and NaCl are also often used although they may be present in the diluent rather than the formulation itself. In addition, compounds such as hydroxyethyl starch can be used to raise the T_c of the product (i.e., to increase T_g'). Especially critical for biopharmaceuticals, stabilizers are often necessary to provide a sufficiently robust formulation. The complex relationship between water content, molecular mobility, and the physical and chemical degradation of dried macromolecular systems can make the selection and optimization of stabilizers especially challenging. There are, however, several generally accepted principles for successful stabilization

of biomolecules in the solid state.140 First, it is well established that the stabilizer must remain amorphous and in the same phase as the drug. Conversely, physical mixtures do not effectively stabilize.141 Second, the stabilizer should be chemically and physically inert. A well-known example of this problem involves the use of sucrose, often a highly effective stabilizer. At low pH, this disaccharide can be hydrolyzed to reducing sugars, which can covalently interact with proteins. Third, as mentioned previously, the formulation should not permit selective buffer crystallization and consequent pH shifts. Many macromolecules and viruses and other biological entities are often pH sensitive with losses in biological properties upon exposure to extremes of pH. Finally, it is very clear that low water content is often essential for optimal stabilization in dried formulations. It should also be mentioned that ice is a major stress during freezing. The formulation of an ice/water interface may result in adsorption of proteins and other amphipathic macromolecules which can be significantly destabilizing. This is at least partially due to the forces exerted by the surface on macromolecules due to the multipoint nature of the contacts between the surface and the drug or vaccine. The presence of stabilizers may reduce such destabilizing effects but the mechanisms are incompletely understood. The preferential hydration (solute exclusion) mechanism discussed previously may be operative at this level. In the solid state, several

factors are considered to play critical roles in stabilization. Chief among these is mobility. It has long been thought that dried formulations are most stable in the glassy, solid state. The existence of such a state, however, does not guarantee long-term stability, even at moderate temperatures. In addition, the presence of the “native” state of a protein is generally considered necessary. This is most frequently analyzed by FTIR spectroscopy which can conveniently analyze protein, nucleic acid, and viral structure in the solid state. A common observation is that without the presence of stabilizers, lyophilized proteins are not in their native state in their solid forms, producing accelerated chemical as well as physical degradation.

A number of mechanisms have been proposed to explain how stabilizers are able to maintain macromolecular structure under conditions of low moisture.140 The water substitution hypothesis argues that many stabilizers interact with proteins and other biological entities in a manner similar to water. This is proposed to support the native state of such molecules and provide stabilization during freezing and drying by providing an appropriate physical environment. Two main lines of evidence in support of this hypothesis are that (*a*) many stabilizers are sugars and due to their multiple hydroxyl groups are able to hydrogen bond to macromolecular systems in a manner similar to water and (*b*) spectroscopic studies demonstrate water-like interactions between stabilizers and biomolecules in the solid state. The water substitute hypothesis has primarily been used to explain stabilization during drying rather than during storage. The second major hypothesis postulates that by creation of a glassy state, there results in a reduction in macromolecular mobility which leads to a decrease in the rate of degradative events. Two types of motions are recognized in glasses. The more global dynamic behavior is known as α-relaxation. It is directly related to viscosity and involves long time and length scales. Conversion to this behavior occurs when solids are converted to liquidlike states and are measured by T_g. Fast dynamic behavior is designated β-relaxation. This involves local motions on a much shorter length and time scale and can be measured by a variety of methods including dielectric, neutron scattering and NMR relaxation techniques. The relationship between the effects of stabilizers on these different types of motions is not simple, however. While it might be expected that stabilizers would simply decrease the amplitude of such processes, both increases and decreases have been seen similar to observations in solution. Furthermore, lowering the T_g does not always destabilize. This remains a very active area of current research with a consensus that dynamics are important, but their precise role yet to be definitively defined.

Some tentative conclusions about the mechanisms of excipient stabilization can, however, be advanced. Cryoprotection may involve solvent exclusion if instability occurs early in freezing. If it occurs later, immobilization by vitrification is more likely. If surfaces are involved, the coating of such surfaces by a surface-active agent such as a surfactant or protein (serum albumin, gelatin, etc.) may be helpful. Lyoprotectants are usually amorphous, chemically inert glass formers. They form single phases with macromolecules and “moderately” interact with their surfaces. They should couple all relevant modes of motion, both local and global, to the matrix and preserve native structure during freezing and drying. The requirements for storage stabilization are similar but specific considerations are usually necessary for each individual biomaterial based on its specific sensitivity to their unique degradation pathways.

Here are a few general rules to guide formulation of macromolecules for freeze-drying. (*a*) The amount of buffer should be minimized (avoid phosphates). (*b*) Employ other salts only if needed (i.e., for solubility) and minimize their amounts. (*c*) Maximize T_g' (<-35°C is usually a problem; <-40°C is typically unacceptable). Lyophilization should be performed below T_g' (or at least below the collapse temperature) in primary drying and below T_g in secondary drying.

If stability problems arise, the following approaches have often proven successful to minimize such difficulties. If the problem occurs during freeze-drying, it should be isolated to either freeze/thaw or freeze dry stability. If the problem is during freezing, the addition of surfactants (Tweens and Pluronics) or high levels of “excluded solutes” (e.g., amino acids, nonreducing carbohydrates, polyethylene glycols) may be useful. If the problem is seen during drying only, the use of nonreducing carbohydrates such as sucrose or trehalose is often successful. If one has a storage stability problem, the moisture level needs to be carefully controlled. It is not unusual, however, for this to be inadequate to completely solve the

problem. In general, the best additional step has been to use a nonreducing carbohydrate like sucrose at neutral

or basic pH (if moisture control is not a problem). At low pH and higher moisture content, trehalose often proves to be a better choice as a storage stabilizer.

Formulation of Vaccines[87,90,142,143]

The formulation of vaccines is complicated by their varied nature (peptide, protein, VLP, virus, DNA, bacteria, polysaccharide) as well as the need for adjuvants[142,143] in their less immunogenic forms. Nevertheless, the procedures employed are very similar to those described above for individual macromolecules. One apparently important difference is the ultimate goal. In the case of biotherapeutics, one wishes to keep immunogenicity to a minimum. In contrast, we desire to maintain an optimal immunogenicity for vaccines. Fortunately, however, both states can be achieved in the same manner, namely, by maintaining a particular structure (typically the native one) both in vivo and during long-term storage. Thus, the immediate relevance of the discussion in the preceding sections should be evident. What about complex vaccines such as attenuated viruses[87]? In such cases, it is still often possible to treat them as physiochemical systems. One must first isolate the virus in purified (i.e., >90%) form. They can then be subjected to the various stresses as described above and analyzed by the same physical and chemical methods. In this case, however, the resultant signals are the sum of the signals from all of the component viral proteins and nucleic acids weighted by the relative amount of each macromolecule and their individual signal intensities. In the case of many viruses, the measured experimental results will primarily reflect viral coat proteins as well as the integrity of the viral particle. If the rate-limiting degradation events are reflected by changes in the properties of components with large contributions to observed signals, the formulation methods described in the previous sections may well be effective. In many cases, however, live attenuated viruses are needed for efficacious vaccines. In fact, these live entities may constitute a very small minority of the total viral particles in a preparation. It is frequently the case, however, that all of the viral particles in such a mixture undergo the destabilizing changes of interest so that physical and chemical methods can still be used in accelerated stability studies. This is because the events that originally inactivated the majority of the viruses are distinct from those that are relevant to long-term stability studies (the initial inactivating events often occur during the cell culture process as is evident from kinetic studies). A second problem is that many vaccines are now used in combination form (measles, mumps, rubella (MMR); diphtheria, tetanus, pertussis (DPT); etc.).[144] In such cases, the individual components must be studied separately and stabilizers identified for each individual component. Mixtures of stabilizers can then be used to address individual problems. Such approaches can also be applied to VLPs, DNA vaccines containing cationic delivery vehicles, and even entire bacterial cells. The latter is again possible if a critical event in bacterial integrity can be detected during accelerated stability testing. Nevertheless, critical physical and/or chemical degradation events may not be detectable in complex vaccines. In such cases, a trial and error (empirical) method must be used. One then usually employs an animal model (most commonly mice) with maintenance of immunogenicity as judged by stimulation of specific antibody levels (less commonly cytokine production). This is usually done by various types of ELISA assays. Selection of potential excipients is based on the principles outlined above with GRAS agents screened as potential stabilizers. In the case of live antigens such as attenuated viruses, cellular responses may also be measured because of the essential nature of the replicative states. Thus, vaccines such as those for measles can have their ability to kill sensitive cells used as a criterion for efficacy using plague assays.

Another complication is the use of adjuvants with weakly immunogenic antigens such as monomeric recombinant proteins. The most commonly used adjuvants are the aluminum salts. Although only effective at enhancing humoral (antibody) responses, their well-established safety and efficacy profiles have resulted in their widespread use. Aluminum salts are usually used in the form of aluminum hydroxide and aluminum phosphate, the former positively charged at neutral pH, the latter negatively charged under these conditions. A general principle in the use of these adjuvants is that antigens must

be adsorbed to the surface of these particulate salts for them to be effective at enhancing immune responses. Thus, the first step in their formulation is to perform binding studies of antigens to their surface (Fig. 21-35). This is simply done by adding incremental amounts of antigens to aluminum salts, incubating for a short period, followed by separation of unbound antigen from antigen/aluminum salt complexes by centrifugation. The amount of unbound antigen is then measured (typically by optical absorbance or a dye-binding method) and the amount of antigen bound determined by subtraction from the amount added. The general rule is that negatively charged antigens bind to the positively charged aluminum hydroxide and positively charged antigens

to the negatively charged aluminum phosphate. Typical formations contain a milligram or less of aluminum with the amount of antigen more variable but typically in the range of a few micrograms to hundreds of micrograms. The charge on the antigen is usually estimated from its isoelectric point.

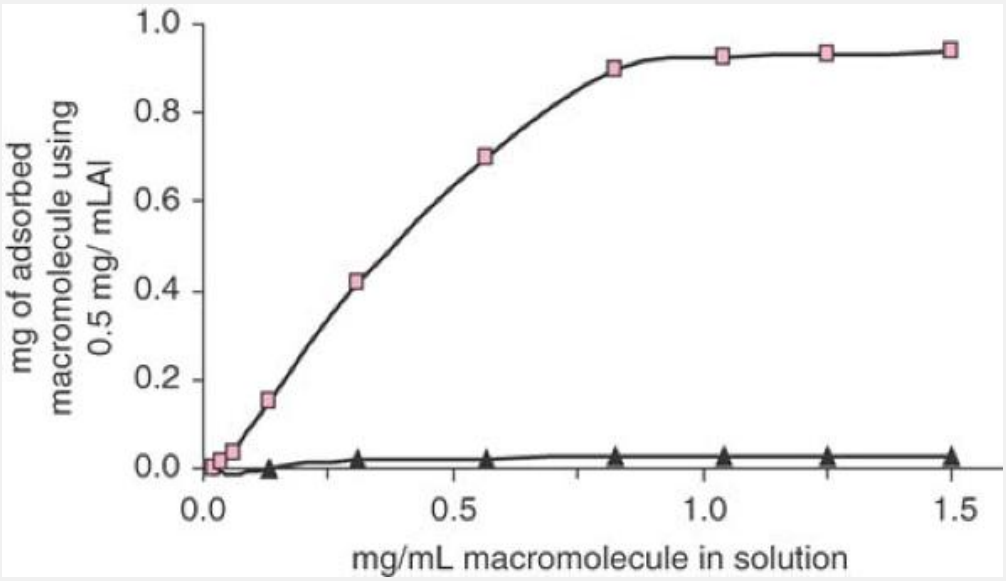

Fig. 21-35. A representative binding isotherm of a vaccine protein to an aluminum salt adjuvant. In this case, the protein is negatively charged (it has a low pI) and binds well to positively charged aluminum hydroxide (squares). In contrast, it binds poorly to negatively charged aluminum phosphate (triangles).

Suspensions of aluminum salts are optically opaque. Thus, methods that can be used to examine adsorbed proteins are more limited than those available for solution studies. Several methods are available, however, to examine proteins and other macromolecular systems adsorbed to aluminum salts adjuvants. For example, fluorescence methods can still be used. Sometimes enough light can still penetrate and exit such suspensions, sufficient for emission spectra to be obtained. If this is not the case, emission can be measured off the surface of the sample by examining emission at a lower angle (e.g., 45°–60°). This is known as front face fluorescence. As discussed earlier, both FTIR (in ATR and DRIFT modes) and Raman spectroscopy can be used in highly scattering samples. Thus, secondary structure information can be obtained by analysis of amide bands by both techniques. The problem of light scattering can be avoided entirely by the use of DSC since this is a thermal technique. The latter methods may not be sensitive enough to detect small amounts of absorbed antigens but fluorescence is usually sensitive enough for such applications. Effects of temperature and pH can be analyzed but it must be shown that the antigen remains adsorbed to the adjuvant surface under these conditions.

Alternatively, the antigen can be removed from the surface and then examined. This can often be done by low pH, high salt concentration, or treatment with an agent that dissolves the adjuvant (e.g., citrate) or a reagent that weakens antigen/ adjuvant interactions such as low concentrations of urea or guanidine hydrochloride (high concentrations cannot be used because they usually disrupt antigen structure). Although the interactions between antigen and adjuvant often contain a major electrostatic component, they may also involve apolar and van der Waal forces among other types of weak noncovalent interactions. A mechanism known as ligand exchange may also occur.145 Unfortunately, it is not uncommon for antigens to be difficult or impossible to remove. This is especially the case after long storage times where the antigen may undergo structural changes as it optimizes its interaction with the aluminum salt surface. In fact, it is often found that antigens are destabilized when they interact with aluminum salts.146 This is typically manifested by a lowering of the T_m measured by DSC, temperature-dependent fluorescence, or vibrational spectroscopic methods. Fortunately, it turns out that compounds that stabilize macromolecules and viruses in solution often also stabilize them on aluminum salt surfaces, although usually to a lesser extent.147 One problem with aluminum salt formulations deserves special mention. Because aluminum hydroxide is positively charged, it attracts hydroxide anions to its surface. This increases surface pH leading to enhanced deamidation of protein antigens. This can often be prevented, however, by the inclusion of small (millimolar) amounts of phosphate in the formulation, which lowers the surface pH by converting small amounts of the aluminum hydroxide to aluminum phosphate. Until recently, it was thought that aluminum salt adjuvants could not be lyophilized. Recent work, however, suggests that the presence of carbohydrate stabilizers will permit this to be done148 and opens up the possibility of using drying technologies to improve the stability of aluminum salt vaccine formulations.

Although the use of other adjuvants is still in its infancy, it is clear that this is unlikely to be the case in the near future. Recent discoveries have found that all mammals possess a series of proteins on the surface of immune cells that recognize highly repetitive structures on the surface of pathogens. These are now known as “toll” receptors.149 Concomitant with these findings, it was determined that many of the adjuvant materials that had been empirically discovered actually acted through binding to these receptors. This is in contrast to aluminum salt adjuvants, which appear to work through a variety of other mechanisms including depot effects, facilitation of antigen entry into cells, and other specific immune effects. A number of toll receptor–based adjuvants have now been tested in human clinical studies. While safety still remains an issue, it seems highly probable that many of these as well as non–toll receptor adjuvants will become available for human use. For example, a number of synthetic lipid A analogues (e.g., monophosphoryl lipid A), saponins, oil-in-water, and water-in-oil emulsions used alone and in combination appear quite promising.142,143 In fact, an oil-in-water emulsion containing squalene and two surfactants (MF59) is already available in a commercial flu vaccine in Europe. Novel adjuvant containing vaccines will present unique formulation problems due to their diversity. It seems probable, however, that the methods currently developed in conjunction with new technologies should be able to meet these challenges.

Chapter Summary

Despite the sophistication of modern biotechnology, significant problems still exist from a pharmaceutical perspective. Like all drugs, biotechnology-based pharmaceuticals and vaccines produce side effects in their recipients. Mechanism-based toxicity as well as adverse effects due to the general physical properties of macromolecules and their complexes remain poorly understood. It is clear that animal models are currently inadequate to address such problems. Improved animal models (especially disease based) are presently an area of great interest as are cell culture systems that might be used to elucidate mechanism-based toxicity. One phenomenon of great concern is the immunogenicity of biopharmaceuticals. This can lead to both loss of activity through neutralization by antibodies as well as pathological immune responses such as allergic reactions. Although immune responses to therapeutic proteins often have little if any negative effects, there is a growing concern that this issue must be more aggressively addressed. Conversely, both recombinant protein and

DNA-based vaccines are usually insufficiently immunogenic. Thus, the development of novel adjuvants as well as improvements in the delivery of DNA vaccines is clearly required. With regard to the latter, nucleic acid–based therapeutics remain inadequately developed and understood, necessitating a greater emphasis on pharmaceutical aspects of their behavior and delivery.

In general, the precise relationships between the structure and behavior of biomolecules in both the solution and solid state are still poorly understood. In many ways, this remains a key to the successful development of biopharmaceuticals from a process, analytical, and formulation perspective. As discussed above, the role of protein dynamics in each of these areas has yet to be definitively explored. One aspect of this poor understanding is a lack of availability of potential stabilizers for use as excipients in biopharmaceutical and vaccine formulations. The GRAS list is, in fact, rather limited and offers a quite restricted number of options to the formulation scientist. As our understanding of biomolecular structure increases, however, we can expect that a combination of rational design and HTP screening methods should allow us to greatly expand stabilizer options after appropriate safety considerations.

A practical problem with biopharmaceuticals is their manufacture at a scale sufficient for use in large populations. For example, in the case of monoclonal antibodies it appears that there is insufficient manufacturing capability if a significant number of such proteins currently in clinical trials come to fruition as marketed pharmaceuticals. The manufacture of live agents such as viruses has always proven to be challenging at an industrial scale. Thus, the development of new technologies to aid in the high-level manufacture of biopharmaceuticals is an important goal of modern biotechnology.

As biotechnology-based products begin to go off patent, the possibility and then the reality of less expensive versions of these drugs and vaccines has become apparent.[150,151]The word “generic” is not generally applied to these agents because they are unlikely to be physically, biologically, and functionally equivalent to the original product. Both the terms “follow-on biologics” and “biosimilars” have been applied to such drugs. The major area of controversy with follow-on recombinant proteins has been the extent to which extensive clinical trials are necessary to ensure their safety and efficacy. In particular, are physical and chemical comparisons of biosimilars to the original innovator drugs sufficient to ensure these critical properties? A recent detailed study comparing the properties of EPO products from a wide variety of different sources[152] emphasizes striking differences based on manufacturing process and company of origin raising significant scientific, legal, and regulatory concerns.

Although at its beginnings, the use of biotechnology (e.g., fermentation, industrial enzymes, etc.) produced little public concern, this changed dramatically when it began to be used to genetically manipulate plants and animals. Scientists themselves expressed concerns that it might be difficult to predict the result of the alteration and insertion of new genes into novel cellular environments. At the public level, this went so far as to imagine the creation of genetically altered organisms with unique pathological characteristics and plants, which might spread deleterious genes into nontarget plants. All of the proceeding can and have occurred, but so far without any significant disasters. Initially, we perhaps forgot that plant and animal breeders have been doing the same thing for hundreds if not thousands of years with essentially positive results. While the potential for problems and negative public perceptions remain real, the success of modern biopharmaceuticals and vaccines has maintained forward momentum in the use of these technologies. Several recent problems, however, illustrate continued negative perceptions in a minority of the population. For example, the use of stem cells derived from embryos to treat various diseases has raised much controversy due to the source of the cells.[153] This is, however, more of an ethical issue rather than a scientific one. Claims that the measles vaccine or the compound thimerosal (a mercury-containing

preservative) in certain vaccines causes autism in children are completely unsubstantiated by scientific evidence but have resulted in significant public concern, nevertheless.[154,155] Such issues will no doubt continue to be raised but do not seem a major impediment to progress in biotechnology.

Practice problems for this chapter can be found at thePoint.lww.com/Sinko6e.

References

1. E. J. Jewell and F. Abate (Eds.), *The New Oxford American Dictionary*, Oxford University Press, New York, 2001.
2. A. D. Smith et al., (Eds.), *Oxford Dictionary of Biochemistry and Molecular Biology*, Rev ed., Oxford University Press, Oxford, 2000.
3. C. Ratledge and B. Kristiansen (Eds.), *Basic Biotechnology*, 3rd Ed., Cambridge University Press, Cambridge, 2006.
4. D. L. Oxender and L. E. Post (Eds.), *Novel Therapeutics from Modern Biotechnology: From Laboratory To Human Testing*, Springer-Verlag, Heidelberg, 1999.
5. W. J. Thieman and M. A. Palladino, *Introduction to Biotechnology*, Benjamin Cummings, San Francisco, 2004.
6. J. E. Smith, *Biotechnology*, 4th Ed., Cambridge University Press, Cambridge, 2004.
7. V. Moses, R. E. Cape, and D. G. Springham (Eds.),*Biotechnology—The Science and The Business*, 2nd Ed., Harwood Academic Publishers, Singapore, 1999.
8. A. L. Demain (Ed.), *Biotechnology for Beginners*, Academic Press, Heidelberg, 2008.
9. G. A. Petsko and D. Ringe, *Protein Structure and Function*, New Science Press Ltd, London, 2004.
10. J. Kyte, *Structure in Protein Science*, Garland Publishing, Inc., New York, 1995.
11. T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd Ed., W. H. Freeman and Company, New York, 1984.
12. S. A. Plotkin, W. A. Orenstein, and P. A. Offit, *Vaccines*, 5th Ed., Saunders, Philadelphia, 2008.
13. B. R. Bloom and P.-H. Lambert (Eds.), *The Vaccine Book*, Academic Press, San Diego, 2003.
14. A. Allen, *Vaccine: The Controversial Story of Medicine's Greatest Lifesaver*, W. W. Norton & Company, New York, 2007.
15. C.-A. Siegrist, in S. A. Plotkin, W. A. Orenstein, and P. A. Offit (Eds.), *Vaccines*, 5th Ed., Saunders, China, 2008, p. 17.

16. R. A. Seder and J. R. Mascola, in B. R. Bloom and P.-H. Lambert (Eds.), *The Vaccine Book*, Academic Press, San Diego, 2003, p. 51.
17. D. V. Schaffer and W. Zhou (Eds.), *Gene Therapy and Gene Delivery Systems (Advances in Biochemical Engineering/Biotechnology)*, Springer-Verlag, Heidelberg, 2006.
18. N. S. Templeton (Ed.), *Gene and Cell Therapy: Therapeutic Mechanisms and Strategies*, 3rd Ed., CRC Press, Boca Raton, 2008.
19. T. A. Cooper, L. Wan, and G. Dreyfuss, Cell, **136,** 777, 2009.
20. D. Castanotto and J. J. Rossi, Nature, **457,** 426, 2009.
21. D. V. Schaffer, J. T. Koerber, and K. I. Lim, Annu. Rev. Biomed Eng. **10,** 169, 2008.
22. D. Schaffert and E. Wagner, Gene Ther. **15,** 1131, 2008.
23. M. A. Mintzer and E. E. Simanek, Chem. Rev. **109,** 259, 2009.
24. J. Kurreck, Angew. Chem. Int. Ed. Engl. **48,** 1378, 2009.
25. K. A. Whitehead, R. Langer, and D. G. Anderson, Nat. Rev. Drug Discov. **8,** 129, 2009.
26. L. Cerchia, P. H. Giangrande, J. O. McNamara, and V. de Franciscis, Methods Mol. Biol. **535,** 59, 2009.
27. G. Mayer, Angew. Chem. Int. Ed. Engl. **48,** 2672, 2009.

28. E. Levy-Nissenbaum, A. F. Radovic-Moreno, A. Z. Wang, R. Langer, and O. C. Farokhzad, Trends Biotechnol. **26,** 442, 2008.
29. H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, and P. Matsudaira, *Molecular Cell Biology*, 6th Ed., W. H. Freeman, New York, 2007.
30. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 5th Ed., Garland Science, New York, 2007.
31. J. Sambrook and D. W. Russell, *Condensed Protocols from Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2006.
32. B. Lewin, *Genes IX*, Jones & Bartlett Publishers, Sudbury, 2007.
33. R. J. Simpson, P. D. Adams, and E. A. Golemis, *Basic Methods in Protein Purification and Analysis: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2008.
34. S. A. Doyle (Ed.), *High Throughput Protein Expression and Purification: Methods and Protocols (Methods in Molecular Biology)*, Humana Press, New York, 2008.
35. J. M. Miller, *Chromatography: Concepts and Contrasts*, 2nd Ed., Wiley-Interscience, New York, 2004.
36. L. Hagel, G. Jagschies, and G. K. Sofer, *Handbook of Process Chromatography: Development, Manufacturing, Validation and Economics*, 2nd Ed., Academic Press, Amsterdam, 2007.
37. N. Salamat-Miller, J. Fang, C. W. Seidel, A. M. Smalter, Y. Assenov, M. Albrecht, and C. R. Middaugh, Mol. Cell Proteomics, **5,** 2263, 2006.
38. N. Salamat-Miller, J. Fang, C. W. Seidel, Y. Assenov, M. Albrecht, and C. R. Middaugh, J. Biol. Chem. **282,** 10153, 2007.
39. K. E. van Holde, W. C. Johnson, and P. S. Ho, *Principles of Physical Biochemistry*, 2nd Ed., Pearson Education Inc., Upper Saddle River, 2006.
40. W. Jiskoot and D. J. A. Crommelin, *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005.
41. S. Peng, C. Cui, M. Zhao, and G. Cui, *Pharmaceutical Bioassays: Methods and Applications*, Wiley, New York, 2009.
42. D. Wild (Ed.), *The Immunoassay Handbook*, 3rd Ed., Elsevier Science, Amsterdam, 2005.
43. J. P. Gosling, *Immunoassays: A Practical Approach*, Oxford University Press, New York, 2000.
44. R. Westermeier, *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations*, 4th Ed., Wiley-VCH, Weinheim, 2005.
45. S. Ahuja and M. Jimidar (Eds.), *Capillary Electrophoresis Methods for Pharmaceutical Analysis: Separation Science and Technology*, **Vol. 9**, Academic Press, San Diego, 2008.
46. W. J. Lough and I. W. Wainer (Eds.), *High Performance Liquid Chromatography: Fundamental Principles and Practice*, Springer, New York, 2008.
47. G. E. Katzenstein, S. A. Vrona, R. J. Wechsler, B. L. Steadman, R. V. Lewis, and C. R. Middaugh, Proc. Natl. Acad. Sci. U.S.A. **83,** 4268, 1986.
48. L. A. Kueltzo and C. R. Middaugh, in W. Jiskoot and D. J. A. Crommelin (Eds.), *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005, p. 1.
49. C. S. Braun, L. A. Kueltzo, and C. R. Middaugh, in M. A. Findeis (Ed.), *Nonviral Vectors for Gene Therapy: Methods and Protocols*, Humana Press, Totowa, 2001, p. 253.
50. L. A. Kueltzo, B. Ersoy, J. P. Ralston, and C. R. Middaugh, J. Pharm. Sci. **92,** 1805, 2003.
51. H. Mach, C. R. Middaugh, and R. V. Lewis, Anal. Biochem.**200,** 74, 1992.
52. C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, and T. Gray, Protein Sci. **4,** 2411, 1995.
53. W. Jiskoot, A. J. W. G. Visser, J. N. Herron, and M. Sutter, in W. Jiskoot and D. J. A. Crommelin (Eds.), *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005, p. 27.
54. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd Ed., Springer, Singapore, 2006.
55. L. H. Lucas, B. A. Ersoy, L. A. Kueltzo, S. B. Joshi, D. T. Brandau, N. Thyagarajapuram, L. J. Peek, and C. R. Middaugh, Protein Sci. **15,** 2228, 2006.
56. N. Berova, K. Nakanishi, and R. W. Woody (Eds.), *Circular Dichroism*, 2nd Ed., Wiley-VCH, New York, 2000.

57. G. D. Fasman (Ed.), *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York, 1996.
58. V. A. Bloomfield, D. M. Crothers, and I. Tinoco, Jr., *Nucleic Acids*, University Science Books, Sausalito, 2000.
59. Z. Ganim, H. S. Chung, A. W. Smith, L. P. Deflores, K. C. Jones, and A. Tokmakoff, Acc. Chem. Res. **41,** 432, 2008.
60. A. Barth, Biochim Biophys. Acta. **1767,** 1073, 2007.
61. M. C. Manning, Expert. Rev. Proteomics **2,** 731, 2005.
62. C. R. Middaugh, H. Mach, J. A. Ryan, G. Sanyal, and D. B. Volkin, in B. A. Shirley (Ed.), *Protein Stability and Folding: Theory and Practice*, Humana Press, Totowa, 1995, p. 137.
63. M. van de Weert, J. A. Hering, and P. I. Haris, in W. Jiskoot and D. J. A. Crommelin (Eds.), *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005, p. 63.
64. Z. Q. Wen, J. Pharm. Sci. **96,** 2861, 2007.
65. G. J. Thomas, Jr., Annu. Rev. Biophys. Biomol. Struct. **28,** 1, 1999.
66. S. Choosakoonkriang, C. M. Wiethoff, G. S. Koe, J. G. Koe, T. J. Anchordoquy, and C. R. Middaugh, J. Pharm. Sci. **92,** 115, 2003.
67. S. Choosakoonkriang, C. M. Wiethoff, T. J. Anchordoquy, G. S. Koe, J. G. Smith, and C. R. Middaugh, J. Biol. Chem. **276,**8037, 2001.
68. B. A. Lobo, S. A. Rogers, S. Choosakoonkriang, J. G. Smith, G. Koe, and C. R. Middaugh, J. Pharm. Sci. **91,** 454, 2002.
69. S. E. Harding, D. B. Sattelle, and V. A. Bloomfield, *Laser Light Scattering in Biochemistry*, Science and Behavior Books, Palo Alto, 1992.
70. B. Chu, *Laser Light Scattering*, Academic Press, New York, 1974.
71. B. J. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*, John Wiley & Sons, Inc., New York, 1976.
72. J. Demeester, S. S. de Smedt, N. N. Sanders, and J. Haustraete, in W. Jiskoot and D. J. A. Crommelin, *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005, p. 245.
73. C. M. Wiethoff and C. R. Middaugh, in M. A. Findeis (Ed.),*Nonviral Vectors for Gene Therapy: Methods and Protocols*, Humana Press, Totowa, 2001, p. 349.
74. J. S. Philo, in W. Jiskoot and D. J. A. Crommelin (Eds.),*Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005, p. 379.
75. J. Liu and S. J. Shire, J. Pharm. Sci. **88,** 1237, 1999.
76. A. Schön and A. Velázquez-Campoy, in W. Jiskoot and D. J. A. Crommelin (Eds.), *Methods for Structural Analysis of Protein Pharmaceuticals*, AAPS Press, Arlington, 2005, p. 573.
77. G. W. H. Höhne, W. F. Hemminger, and H.-J. Flammersheim, *Differential Scanning Calorimetry*, 2nd rev ed., Springer, Berlin, 2003.
78. S. Gaisford and M. A. A. O'Neill, *Pharmaceutical Isothermal Calorimetry*, Informa Health Care, New York, 2006.
79. J. E. Ladbury and M. L. Doyle (Eds.), *Biocalorimetry 2*, John Wiley & Sons, Ltd., Chichester, 2004.
80. C. H. Spink, Methods Cell Biol. **84,** 115, 2008.

81. P. L. Privalov and A. I. Dragan, Biophys. Chem. **126,** 16, 2007.
82. B. A. Lobo, S. A. Rogers, C. M. Wiethoff, S. Choosakoonkriang, S. Bogdanowich-Knipp, and C. R. Middaugh, in M. A. Findeis (Ed.), *Nonviral Vectors for Gene Therapy: Methods and Protocols*, Humana Press, Totowa, 2001, p. 319.
83. B. A. Lobo, G. S. Koe, J. G. Koe, and C. R. Middaugh, Biophys. Chem. **104,** 67, 2003.
84. T. J. Kamerzell and C. R. Middaugh, J. Pharm. Sci. **97,**3494, 2008.

85. V. J. Hilser, E. B. Garcia-Moreno, T. G. Oas, G. Kapp, and S. T. Whitten, Chem. Rev. **106,** 1545, 2006.
86. R. Esfandiary, J. S. Hungan, G. H. Lushington, S. B. Joshi, and C. R. Middaugh, Protein Sci., In Press, 2009.
87. C. J. Burke, T. A. Hsu, and D. B. Volkin, Crit. Rev. Ther. Drug Carrier Syst. **16,** 1, 1999.
88. D. B. Volkin, G. Sanyal, C. J. Burke, and C. R. Middaugh, Pharm. Biotechnol. **14,** 1, 2002.
89. J. F. Carpenter and M. C. Manning (Eds.), *Rational Design of Stable Protein Formulations: Theory and Practice (Pharmaceutical Biotechnology)*, Springer, New York, 2002.
90. D. T. Brandau, L. S. Jones, C. M. Wiethoff, J. Rexroad, and C. R. Middaugh, J. Pharm. Sci. **92,** 218, 2003.
91. K. A. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*, Garland Science, New York, 2003.
92. A. L. Fink, in B. A. Shirley (Ed.), *Protein Stability and Folding: Theory and Practice*, Humana Press, Totowa, 1995, p. 343.
93. B. A. Shirley, in B. A. Shirley (Ed.), *Protein Stability and Folding: Theory and Practice*, Humana Press, Totowa, 1995, p. 177.
94. C. N. Pace and K. L. Shaw, Proteins, Suppl 4, 1, 2000.
95. J. K. Myers, C. N. Pace, and J. M. Scholtz, Protein Sci. **4,**2138, 1995.
96. G. R. Grimsley, J. M. Scholtz, and C. N. Pace, Protein Sci.**18,** 247, 2009.
97. C. N. Pace, G. R. Grimsley, and J. M. Scholtz, J. Biol. Chem.**284,** 13285, 2009.
98. C. R. Middaugh, R. K. Evans, D. L. Montgomery, and D. R. Casimiro, J. Pharm. Sci. **87,** 130, 1998.
99. T. J. Ahern and M. C. Manning (Eds.), *Stability of Protein Pharmaceuticals*, Plenum Press, New York, 1992.
100. D. Pogocki and C. Schöneich, J. Pharm. Sci. **89,** 443, 2000.
101. S. Clarke, Int. J. Pept. Protein Res. **30,** 808, 1987.
102. K. Patel and R. T. Borchardt, J. Parenter Sci. Technol. **44,**300, 1990.
103. K. Patel and R. T. Borchardt, Pharm. Res. **7,** 787, 1990.
104. K. Patel and R. T. Borchardt, Pharm. Res. **7,** 703, 1990.
105. C. L. Stevenson, M. E. Donlan, A. R. Friedman, and R. T. Borchardt, Int. J. Pept. Protein Res. **42,** 24, 1993.
106. C. L. Stevenson, A. R. Friedman, T. M. Kubiak, M. E. Donlan, and R. T. Borchardt, Int. J. Pept. Protein Res. **42,** 497, 1993.
107. M. Xie, J. Aube, R. T. Borchardt, M. Morton, E. M. Topp, D. Vander Velde, and R. L. Schowen, J. Pept. Res. **56,** 165, 2000.
108. N. E. Robinson and A. B. Robinson, Proc. Natl. Acad. Sci. U.S.A. **98,** 4367, 2001.
109. C. Oliyai and R. T. Borchardt, Pharm. Res. **10,** 95, 1993.
110. C. Oliyai and R. T. Borchardt, Pharm. Res. **11,** 751, 1994.
111. C. Oliyai, J. P. Patel, L. Carr, and R. T. Borchardt, Pharm. Res. **11,** 901, 1994.
112. S. J. Bogdanowich-Knipp, S. Chakrabarti, T. D. Williams, R. K. Dillman, and T. J. Siahaan. J. Pept. Res. **53,** 530, 1999.
113. S. J. Bogdanowich-Knipp, D. S. Jois, and T. J. Siahaan, J. Pept. Res. **53,** 523, 1999.
114. Y. Sohma, Y. Hayashi, M. Skwarczynski, Y. Hamada, M. Sasaki, T. Kimura, and Y. Kiso, Biopolymers, **76,** 344, 2004.
115. H. T. He, R. N. Gursoy, L. Kupczyk-Subotkowska, J. Tian, T. Williams, and T. J. Siahaan, J. Pharm. Sci. **95,** 2222, 2006.
116. D. B. Volkin and A. M. Klibanov, J. Biol. Chem. **262,** 2945, 1987.
117. S. Li, C. Schöneich, and R. T. Borchardt, Biotechnol. Bioeng. **48,** 490, 1995.
118. S. Li, T. H. Nguyen, S. Schöneich, and R. T. Borchardt, Biochemistry, **34,** 5762, 1995.
119. S. Li, C. Schöneich, and R. T. Borchardt, Pharm. Res. **12,**348, 1995.
120. S. Li, C. Schöneich, G. S. Wilson, and R. T. Borchardt, Pharm. Res. **10,** 1572, 1993.

121. C. Schöneich, F. Zhao, G. S. Wilson, and R. T. Borchardt, Biochim. Biophys. Acta. **1158,** 307, 1993.
122. K. Uchida and S. Kawasaki, Biochem Biophys. Res. Commun. **138,** 659, 1986.
123. T. J. Simat and H. Steinhart, J. Agric. Food Chem. **46,** 490, 1998.
124. K. Itakura, K. Uchida, and S. Kawakishi, Chem. Res. Toxicol. **7,** 185, 1994.
125. E. L. Finley, J. Dillon, R. K. Crouch, and K. L. Schey, Protein Sci. **7,** 2391, 1998.
126. C. Giulivi, N. J. Traaseth, and K. J. Davies, Amino Acids,**25,** 227, 2003.
127. A. Beck, M. C. Bussat, C. Klinguer-Hamour, L. Goetsch, J. P. Aubry, T. Champion, E. Julien, J. F. Haeuw, J. Y. Bonnefoy, and N. Corvaia, J. Pept. Res. **57,** 528, 2001.
128. S. N. Timasheff, Biochemistry, **41,** 13473, 2002.
129. H. Fan, S. N. Vitharana, T. Chen, D. O'Keefe, and C. R. Middaugh, Mol. Pharm. **4,** 232, 2007.
130. T. J. Kamerzell, S. B. Joshi, D. McClean, L. Peplinskie, K. Toney, D. Papac, M. Li, and C. R. Middaugh, Protein Sci. **16,**1193, 2007.
131. S. F. Ausar, M. Espina, J. Brock, N. Thyagarajapuram, R. Repetto, L. Khandke, and C. R. Middaugh, Hum. Vaccin. **3,** 94, 2007.
132. C. J. Burke, B. L. Steadman, D. B. Volkin, p. K. Tsai, M. W. Bruner, and C. R. Middaugh, Int. J. Pharm. **86,** 89, 1992.
133. H. R. Costantino and M. J. Pikal, *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004.
134. B. S. Chang and S. Y. Patro, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 113.
135. Y.-F. Maa and H. R. Costantino, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 519.
136. C. A. Angell and J. L. Green, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 367.
137. D. E. Overcashier, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 337.
138. D. Lechuga-Ballesteros, D. P. Miller, and S. P. Duddu, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 271.
139. H. R. Costantino, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 139.
140. J. F. Carpenter, B. S. Chang, and T. W. Randolph, in H. R. Costantino and M. J. Pikal (Eds.), *Lyohilization of Biopharmaceuticals*, AAPS Press, Arlington, 2004, p. 423.
141. P. O. Souillac, H. R. Costantino, C. R. Middaugh, and J. H. Rytting, J. Pharm. Sci. **91,** 206, 2002.
142. C. J. Hackett and D. A. Harn, Jr. (Eds.), *Vaccine Adjuvants: Immunological and Clinical Principles*, Humana Press, Totowa, 2006.
143. M. Singh (Ed.), *Vaccine Adjuvants and Delivery Systems*, John Wiley & Sons, Inc., Hoboken, 2007.
144. M. D. Decker, K. M. Edwards, and H. H. Bogaerts, in S. A. Plotkin, W. A. Orenstein, and P. A. Offit (Eds.), *Vaccines*, 5th Ed., Saunders, Philadelphia, 2008, p. 1069.
145. F. R. Vogel and S. L. Hem, in S. A. Plotkin, W. A. Orenstein, and P. A. Offit (Eds.), *Vaccines*, 5th Ed., Saunders, Philadelphia, 2008, p. 59.
146. L. S. Jones, L. J. Peek, J. Power, A. Markham, B. Yazzie, and C. R. Middaugh, J. Biol. Chem. **280,** 13406, 2005.
147. L. J. Peek, T. T. Martin, C. Elk Nation, S. A. Pegram, and C. R. Middaugh, J. Pharm. Sci. **96,** 547, 2007.
148. A. L. Clausi, S. A. Merkley, J. F. Carpenter, and T. W. Randolph, J. Pharm. Sci. **97,** 2049, 2008.
149. B. Beutler, in C. J. Hackett and D. A. Harn, Jr. (Eds.),*Vaccine Adjuvants: Immunological and Clinical Principles*, Humana Press, Totowa, 2006, p. 1.
150. D. M. Dudzinski and A. S. Kesselheim, N. Engl. J. Med.**358,** 843, 2008.

151. J. Woodcock, J. Griffin, R. Behrman, B. Cherney, T. Crescenzi, B. Fraser, D. Hixon, C. Joneckis, S. Kozlowski, A. Rosenberg, L. Schrager, E. Shacter, R. Temple, K. Webber, and H. Winkle, Nat. Rev. Drug Discov. **6,** 437, 2007.

152. S. S. Parks, J. Park, J. Ko, L. Chen, D. Meriage, J. Crouse-Zeineddini, W. Wong, and B. A. Kerwin, J. Pharm. Sci. **98,** 1688, 2009.
153. J. M. Wilson, Science, **324,** 727, 2009.
154. P. A. Offit, *Autism's False Prophets: Bad Science, Risky Medicine, and the Search for a Cure*, Columbia Press, New York, 2008.
155. G. A. Poland, R. M. Jacobson, and I. G. Ovsyannikova, Vaccine, **27,** 3240, 2009.

Recommended Reading

It is recommended that the interested student view three excellent films, which provide a nice overview of the history of biotechnology. These are “Glory Enough for All” (the discovery of insulin), “Double Helix” (the discovery of the structure of DNA), and “And the Band Played On” (the early days of the AIDS epidemic). In dramatic form, these three films well illustrate the promise and problems of biotechnology.

Chapter Legacy

Sixth Edition: published as Chapter 21 (Pharmaceutical Biotechnology). This is a new chapter written by Charles Russell Middaugh and Teruna J. Siahaan.