DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 16(18), 2635-2654 (1990)

BIOPHYSICAL APPROACHES TO THE PHARMACEUTICAL DEVELOPMENT OF PROTEINS

C. Russell Middaugh
Merck Sharp & Dohme Research Laboratories
WP26-331
West Point, PA 19486

ABSTRACT

The complexity of protein pharmaceuticals necessitates a multifaceted approach to their characterization, stabilization, and development. The combined use of spectroscopic, hydrodynamic, chromatographic and thermodynamic methods to evaluate various levels of macromolecular structure is described. Near future developments in our ability to evaluate proteins at high structural resolution are also considered.

DISCUSSION

The emerging use of proteins produced by recombinant DNA technologies as pharmaceuticals has resulted in an increased interest in the characterization of these large molecules. If possible, one must demonstrate the structural identity of recombinantly manufactured proteins and their natural homologues and/or the maintenance of the structural and

biological integrity of the genetically engineered molecules throughout their lifetime. Unlike lower molecular weight compounds whose identity and correctness of structure can reliably be established by chromatographic behavior and one or a few simple structural determinations (e.g. NMR, FTIR, MS), much more elaborate procedures are currently necessary to accomplish this task with proteins. This additional effort arises directly from the intrinsic structural complexity of peptide macromolecules.

Proteins are encountered in a wide variety of sizes and shapes. They range in molecular weight from a few thousand Daltons into the millions. While many possess roughly globular shapes, they are also found in elongated forms including fibers as well as in more defined configurations such as the pentameric spider-like immunoglobulin M molecule and the hydra headed first component of complement, C1q. They also display considerable microheterogenity, further complicating their description. Related to this structural complexity is the marginal conformational stability of proteins which are maintained in their unique three dimensional form by usually only 2-15 Kcal of free energy. The specific bioactive form of a protein is consequently very sensitive to solution conditions of temperature, pH, ionic strength and the presence of specific solution components. The conditions under which a protein is optimally stable must usually be determined empirically. As our understanding of the biological environment in which a

particular protein evolved improves, we may receive some guidance in this regard. This is especially the case for situations of high thermodynamic activity since the presence of other macromolecular structures in vivo produces large excluded volume effects which in turn lead to large activity coefficients for individual macromolecules.

The fine structural details of proteins lead directly to their exquisite specificity and high biological potency. This is well illustrated by the fine structure of antibody combining sites. Crystal structures of several immunoglobulins clearly show a highly convoluted, intricate surface architecture formed from hypervariable regions capable of unique, multipoint interaction with antigen. The addition or loss of a single contact point between antigen and antibody can dramatically alter specificity. A consequence of high specificity and potency is the requirement for low concentrations of protein, necessitating particularly rigorous, analytical methodologies.

Owing to the normal complexity of protein structure, it is convenient to recognize various levels of such structure, both for conceptual and analytical purposes. The primary structure of a protein is the linear order of the amino acids that constitute its polypeptide chain(s). A complete description of a protein's primary (chemical, covalent) structure may also include identification of glycosylation or other modification sites (e.g. sulfation, phosphorylation) as well as a description of the modifying entities themselves.

For simplicity, we will omit the important considerations of glycosylation and other covalent additions and focus on the polypeptide portion of the protein molecule.

Although it is sometimes stated that the sequence of a protein directly determines higher order structure, this is usually an over-simplification especially for analytical purposes. A protein's secondary structure is defined as the structure that forms as a result of local interaction within a polypeptide chain. Most commonly we recognize α-helices, ß-sheets and turn regions in this category with remaining structure grouped into an "other" or "disordered" class. It has recently become convenient to recognize a "super-secondary" type of structure describing assemblies of secondary structure elements by their common appearance (e.g. Greek-key motifs, ß-barrels, etc.), but this subdivision will not be considered here because of the relatively experimentally inaccessible nature of such structures. Tertiary structure is considered to include those aspects of chain folding that result from more distant interactions within the polypeptide chain. A complete three-dimensional description of a protein molecule is essentially a description of its tertiary structure. Quaternary structure reflects interactions between polypeptides and most commonly can be equated with the subunit structure of a protein. Finally, it should be recognized that a fifth level of structure description may eventually play an important role

in the overall description of proteins. This is the dynamic structure of a protein molecule. It is now clear that proteins exist in a large number of structurally distinct conformations undergoing rapid interconversion at solution temperatures. The motions of the protein responsible for these effects range in scale from simple rotations of side chains to larger cooperative movements of regions of polypeptide chains (i.e. breathing modes). An example with known physiological significance is the movement of the Fab arms of immunoglobulins which facilitates antigen crosslinking.

Given the hierarchical nature of protein structure, two approaches exist to test for the proper structure of a protein. Ideally one would employ a technique that monitors several aspects of protein structure simultaneously. Sufficient information should be contained in any such measurement that structural changes of virtually any type and magnitude are detectable. In principle, two such techniques exist, but neither is currently readily applicable to the pharmaceutical analysis of proteins. The first is X-ray crystallography. Although this approach gives detailed, high resolution structural information about a protein, it requires highly ordered crystals, large amounts of material, and cannot be applied under solution conditions. Thus, it seems unlikely that crystallography will be directly employed as a tool for the pharmaceutical development of proteins in the near future.

The second method that might be employed is nuclear magnetic resonance (NMR) in a two or higher dimensional format. This method is directly applicable to proteins in solution and has the potential to supply high resolution structural information. Unfortunately, NMR suffers from two critical limitations. It requires high concentrations of protein (typically millimolar) and can only be applied to small proteins (less than twenty thousand molecular weight).

Because of these problems, crystallography and NMR are of use only under very special conditions and an entirely different approach to structural analysis of proteins needs to be under- taken during the development of protein drugs. Rather than attempt to obtain a multitude of structural information in a single measurement, a series of measurements are performed that focus on the various individual elements of protein structure. Ideally, such techniques should fulfill the following criteria: they should be applicable over a wide range of protein concentration; they should permit measurement under varying solution conditions (of, e.g., pH and temperature); the presence of excipients should not impede measurements; and little or no sample preparation should be required. While not all the methods to be described meet these requirements, the use of multiple techniques with overlapping sensitivity to the various features of protein structure has the potential to build a comprehensive picture of a protein under

pharmaceutically relevant conditions. These methods will now be briefly examined in terms of their ability to probe the various elements of protein structure.

The primary structure of a protein can be directly determined by automated gas phase sequencing methods. Recently, mass spectroscopy has been used to obtain sequence data, but instrumentation for this purpose currently has limited availability. Both of these procedures are very labor intensive and are not routinely undertaken on protein samples during the development process, but rather are performed only on the bulk protein material after manufacture. Most commonly, peptide mapping is done to ensure sequence integrity. In this procedure, protein is chemically or enzymatically cleaved and fragments chromatographically separated. The elution behavior of the individual fragments can usually be sufficiently optimized to detect alterations in elution position due to covalent changes. This method can often be performed under a variety of solution conditions (characterization is required for each) and is rapid and reproducible. The correctness of a protein's sequence can be less directly examined by chromatography of the intact macromolecule. Electrophoretic methods (e.g. sodium dodecyl sulfate polyacrylamide gel [SDS-PAGE], capillary zone), HPLC (reversed-phase and ion-exchange), and isoelectric focusing are often used for this purpose. The sensitivity of these methods, especially when

used in combination, is often surprisingly good but great care must be taken to show that they can detect the specific changes that can occur within a particular protein.

The present method of choice for analyzing the secondary structure of a protein is circular dichroism (CD). In this method, the unique optical activity of peptide bonds in regular secondary structure in the far ultraviolet (185-250nm) is monitored by measuring the difference in absorption of left and right handed circularly polarized light. The major limitation of CD is interference by other agents in the UV region. A fairly wide range of protein concentrations can be examined by judicious choice of optical path length. Absolute amounts of secondary structure can be estimated by reference to spectra of proteins of known secondary structure content, but the primary value of this technique in the pharmaceutical analysis of proteins is in searching for changes in secondary structure, an application for which its sensitivity is in the one to three percent range.

The recent advent of Fourier transform infrared spectroscopy (FTIR) has provided another convenient method to monitor protein secondary structure. The Amide I region (1600-1700 cm^{-1}) of a protein's infrared spectrum contains signals characteristic of the various types of secondary structure. Spectra can now be obtained in water (although more easily in deuterium oxide) over a wide range of protein

concentrations (in some cases down to 0.1 mg/ml) and in a variety of physical states, including solids when sampling methods such as diffuse and internal reflectance are employed. The use of Fourier self deconvolution and second derivative data analysis permit secondary structure content to be estimated, but like CD, this procedure is best utilized in a comparative mode. Raman spectroscopy can also be employed in a similar manner, but is limited by the very high concentrations usually needed to obtain spectra of sufficient resolution and by the occasional problem of background fluorescence.

In principle, information about protein secondary structure is also present in the far UV absorption spectrum, but frequent lack of suitable instrumentation (monochrometers must extend below 190nm and nitrogen flushing is required) and interference from solvent absorption has so far limited extensive application of this technology. As indicated above, 2D-NMR spectra also contain secondary structure information in the form of characteristic patterns of cross peaks, but attempts to employ this method have so far been inhibited by problems of protein size and concentration.

Comprehensive analysis of protein tertiary structure can currently only be performed by crystallography and NMR. Several methods exist, however, that can provide useful information about certain regions of protein molecules. Many of these methods focus on the aromatic residues of proteins. Thus, only

changes in structure in the vicinity of tryptophan, tyrosine and phenylalanine side chains are probed.

Furthermore, in some cases only combined effects are measured, since multiple aromatic residues with overlapping spectra are present in most proteins. To our advantage, however, the presence of significant quantities of such amino acids in proteins may permit quite subtle alterations in structure to be detected. The easily obtained near UV absorbance spectra of proteins contain contributions from all three aromatic sidechains. Spectral properties of individual residues can now be easily resolved by derivative spectroscopy, making this method of analysis one of increasing utility, especially with the introduction of diode array chromatographic detection systems. Intrinsic fluorescence spectra of proteins are usually dominated by emission from indole sidechains, if tryptophan is present in the molecule. The emission properties of the indole ring are very sensitive to the polarity of their micro-environment, making this approach generally much more sensitive than simple absorbance spectroscopy. Furthermore, the ability to measure lifetimes of excited states, the polarization and intensity of emitted light, as well as excitation spectra make fluorescence spectroscopy a tool of great versatility. The near UV CD spectra of proteins also contain a complex series of overlapping peaks due to optical asymmetry in the environment of aromatic residues. These

spectral features are frequently sensitive to their environment, but high protein concentrations are usually required for such measurements. The Raman spectra of proteins contain peaks from tryptophan and tyrosine residues as well as disulfides that are conformationally sensitive, but very high protein concentrations are again necessary to obtain useful spectra.

The tertiary structure of a protein can also be characterized in terms of the overall shape of the molecule. Several different methods are available for this type of analysis. Perhaps the most powerful approach currently employed is a combination of static and dynamic light scattering. Measurement of the intensity of scattered light as a function of the scattering angle allows calculation of the molecular weight and radius of gyration of a molecule if the dependence of the particle's refractive index on its concentration is known. In a much simpler and more convenient experiment, the Stoke's (hydrodynamic) radius of a particle and a measure of its polydispersity can be determined from an analysis of the time-dependent fluctuations in intensity of the scattered light. Both static and dynamic measurements can be made simultaneously and comparison of the radius of gyration with the hydrodynamic radius gives information about the topological asymmetry of the molecule.

Another method to characterize size and shape involves analyzing the effect of a centrifugal field on a macromolecule. Analysis of the velocity of the particle can give the molecular weight if the particle's partial specific volume and diffusion coefficient are known. The sedimentation coefficient, which can be directly derived from such an experiment, is often very sensitive to the shape of the sedimenting entity. If the particle is sedimented to equilibrium, measurement of the equilibrium distribution of the particle's concentration at different points in the experimental cell combined with the partial specific volume can also give an unambiguous determination of molecular weight. Both methods are particularly useful in the analysis of association/dissociation equilibria of proteins. The recent reintroduction of a commercially available analytical ultracentrifuge promises to make this once again a frequently used technique.

Probably the most common and simplest way of detecting overall shape changes in proteins employs size exclusion (molecular sieve, gel filtration) chromatography. The use of this technique in an HPLC mode has greatly increased its utility by both increasing the speed of analysis and reducing the required sample size. Although it is not particularly sensitive to small conformational changes and relies on comparison to standards for quantitative determination of molecular weight or hydrodynamic size, its convenience and

simplicity often make it the method of first choice when attempting to detect shape changes. Finally, differential scanning calorimetry (DSC) can be used to detect the presence of domains in proteins by measuring the heat absorption associated with the temperature induced unfolding of the structural elements. Independent or partially independent unfolding of domains may be manifested as complexity in the unfolding endotherms which can be analyzed in terms of the integrity of each domain.

Quaternary structure is usually examined by comparing the size of a protein under native and denaturing conditions. Methods such as electrophoresis (e.g. SDS-PAGE), light scattering, sedimentation, size exclusion chromatography and fluorescence polarization have all been employed for this purpose. This approach is often supplemented with chemical crosslinking using bifunctional reagents to stabilize native and intermediate forms of oligomeric proteins. It is sometimes possible to use mass action to dissociate oligomeric proteins into their constituent subunits by going to very low protein concentrations. Any of the above methods can then be employed to determine size if it has sufficient sensitivity under the experimental conditions. An infrequently used but sensitive approach is to determine the partitioning of proteins in aqueous two-phase polymer systems (e.g. PEG and dextran) as a function of protein concentration. If the partitioning of the

various forms of the complex varies, it is sometimes possible to characterize quaternary structure from concentration-dependent changes in the protein's partition coefficient using this technique.

The dynamic structure of proteins has been much less frequently characterized in pharmaceutical studies of proteins, but several straightforward types of analyses can be readily performed. A protein's amides can be labeled at low pH with tritium or deuterium and the out-exchange of the isotope followed at neutral pH by gel filtration followed by radioactivity determination or FTIR (Amide II band), respectively. Penetration of small molecules into the interior of proteins, a process presumably mediated by fluctuations in protein structure, can be followed by the quenching of the fluorescence of buried indole moieties by specific solutes. Oxygen, iodide and acrylamide are most often used for this purpose. Numerous other methods such as NMR and Raman spectroscopy, and determination of distributions of fluorescence lifetimes are also used to probe a range of structural motion times, but are unlikely to be useful for routine pharmaceutical applications due to experimental limitations.

Given the extensive nature of the characterization of a protein that can be performed using the above described plethora of techniques, do we then have the ability to establish the structural correctness of a protein sufficient to

ensure its biological activity and subsequent clinical efficacy? Unfortunately, the answer at present must be a definite no. It is clear that small structural alterations capable of causing an alteration in the activity of a protein could pass by the vast array of indicated biophysical methods undetected (with the probable exception of NMR; see below). In contrast, quite satisfactory analyses of structural stability can be undertaken using these methods. Furthermore, small changes in structure not seen under fixed environmental conditions by biophysical techniques can sometimes be detected by stressing the macromolecule, i.e., small structural alterations in proteins may be manifested as large changes in stability.

Common changes in the covalent structure of proteins include hydrolysis (e.g. of peptide bonds), disulfide exchange, oxidation (e.g. of cysteines and methionines), deamidation (of asparagine and glutamine residues), photoinduced alterations (especially of tryptophan) and racemizations. Such changes can be detected by the methods sensitive to primary structure discussed previously such as amino acid sequencing, peptide mapping and various chromatographic methods including reversed phase HPLC and most recently capillary electrophoresis. Changes in noncovalent structure range from various small conformational changes in highly localized regions of a polypeptide to complete disruption of tertiary structure

(denaturation). These alterations may lead in turn to aggregation or dissociation of oligomeric species and ultimately precipitation in the former case. Conformational changes are most often detected by spectroscopic methods (CD, fluorescence, etc.) while association phenomena are usually easily seen by size exclusion chromatography, light scattering and sedimentation studies.

Ultimately, it is necessary to examine protein stability under pharmaceutical and clinical conditions of storage and use. All of the techniques indicated above are useful for this purpose. The very long incubation times required for such experiments, however, suggest the use of accelerated stability conditions to obtain a preliminary indication of a protein's stability. High temperature is most often used for this purpose. Temperatures which either slowly (seconds to minutes) or rapidly (hours to days) produce structural changes can be selected based on preliminary studies. Temperature induced structural changes are usually easily monitored by CD, fluorescence, UV absorption, DSC or light scattering (often in the form of simple turbidity measurements). Using the same methods, the effect of pH can also be ascertained. Solutes can also be used to disrupt the structure of proteins. Frequently employed in this regard are amides such as urea and guanidine hydrochloride, chaotropic salts, detergents and organic solvents. Protein specific agents are applicable in limited

cases (e.g. reducing or oxidizing agents, metal chelators). Solute addition is generally less satisfactory than extremes of temperature and pH since the complexity of their mode of action makes extrapolation to conditions of interest somewhat tenuous.

Accelerated studies are often used to screen for excipients that increase the stability of a particular protein. The science of protein formulation has so far been primarily empirical. Commonly used additives include amino acids, sugars, surfactants, redox reagents, metals, polyanions and other proteins (to minimize surface adsorption). More protein specific reagents can be expected to be used in the future as the type of biophysical studies discussed increase in number and sophistication.

To the pharmaceutical scientist encountering proteins for the first time, the type of comprehensive biophysical analysis presented above may appear somewhat daunting. Despite the analysis of a protein by perhaps ten to twenty different approaches, one is still ultimately left with reliance upon imprecise and often inaccurate bioassays for final structural confirmation. Does the near future hold any hope for improvement upon this situation? The answer is probably, but by no means certainly, yes. Ideally, a single method that is sensitive to all aspects of protein structure could replace the multifaceted approach currently employed. There is little doubt that the most likely candidate for such an approach is

NMR. Although currently limited by molecular size and sensitivity, continuous improvement in both the sensitivity and availability of large NMR spectrometers and in the methodology of data acquisition (multidimensionality, novel pulse sequences, etc.) promise increasing pharmaceutical applications. A similar situation exists for mass spectroscopy, although the amount of higher order structural information that can be obtained from this technique will be limited by the necessity to ionize samples. Other new high resolution techniques such as scanning tunneling and atomic force microscopy may also offer the type of structural resolution needed, although problems of sample preparation and image interpretation currently limit their use.

Alternatively, many techniques may be automated and combined in such a way that extensive structural information can be obtained from a single experiment. For example, one can envision such a system containing a chromatographic column on the front end based on an ion-exchange separation (HPLC or capillary zone electrophoresis). This would be followed by a series of detectors monitoring UV absorbance (information on concentration and tertiary structure and protein electrostatic properties from elution position/mobility), fluorescence (tertiary structure), CD (secondary structure), light scattering (tertiary/quaternary structure) and, finally, mass spectroscopy (covalent structure). A multiparametric analysis

of the data from the detectors especially by reference to a standard could provide sufficient information to ensure structural identity. Conditions for each protein would be established in a preformulation analysis. Further information could be obtained by a second run through the system under optimized stress conditions induced by temperature, pH or solute. Such a system is under current development in several laboratories and will probably require several years of validation before it will be of general use.

The crucial unanswered question is the exact type and amount of information necessary to define the structure of a protein sufficient to ensure biological activity within acceptable limits. The answer must originate from a complex combination of theoretical, experimental, practical and regulatory considerations which have yet to be rigorously performed. It seems even within the realm of current technology, however, that it will be possible to establish such criteria and perhaps limit the use of laborious and imprecise bioassays to final verification of the activity of critical protein pharmaceutical formulations.

RECOMMENDED BIBLIOGRAPHY

C.R. Cantor and P.R. Schimmel, "Biophysical Chemistry Part II. Techniques for the Study of Biological Structure and Function," Freeman, San Francisco , 1980.

T.E. Creighton, "Proteins, Structures and Molecular Properties," Freeman, New York, 1983.

J. Geigert, J. Parent. Sci. Tech., 43, 220 (1989).

H.A. Havel, R.S. Chao, R.J. Haskell and T.J. Thamann, Anal. Chem., 61, 642 (1989).

M.C. Manning, K. Patel and R.T. Borchardt, Pharm. Res., 6, 903 (1989).

Y.-C.J. Wang and M.A. Hanson, J. Parent Sci. Tech., 42, 53 (1989).

A series of volumes entitled "Pharmaceutical Biotechnology" will be published in 1991 by Plenum Press and should serve as a valuable source of information about protein pharmaceuticals.