Biophysical Characterization of Proteins in the Post-genomic Era of Proteomics*

Kenneth E. Neet‡§ and J. Ching Lee¶

Proteomics focuses on the high throughput study of the expression, structure, interactions, and, to some extent, function of large numbers of proteins. A true understanding of the functioning of a living cell also requires a quantitative description of the stoichiometry, kinetics, and energetics of each protein complex in a cellular pathway. Classical molecular biophysical studies contribute to understanding of these detailed properties of proteins on a smaller scale than does proteomics in that individual proteins are usually studied. This perspective article deals with the role of biophysical methods in the study of proteins in the proteomic era. Several important physical biochemical methods are discussed briefly and critiqued from the standpoint of information content and data acquisition. The focus is on conformational changes and macromolecular assembly, the utility of dynamic and static structural data, and the necessity to combine experimental approaches to obtain a full functional description. The conclusions are that biophysical information on proteins is a useful adjunct to “standard” proteomic methods, that data can be obtained by high throughput technology in some instances, but that hypothesis-driven experimentation may frequently be required. *Molecular & Cellular Proteomics* 1:415–420, 2002.

GENOMICS, PROTEOMICS, AND PHYSICAL BIOCHEMISTRY

The genomic era of biomedical research in the 1990s provided massive amounts of information on DNA sequences from many species culminating in the nearly completed human genome sequence (1, 2). This wealth of data has been annotated and is continuing to be analyzed by new bioinformatic algorithms. Many protein sequences have been inferred from open reading frames from the more than 50 complete genome sequences (www.tigr.org/tdb/tgi) that include human, rodents, bacteria, viruses, and plants. The next two experimental steps are well underway in the 21st century. The aim of proteomics is to determine the structure, function, and expression of all proteins and their isoforms of a genome. The aim of structural genomics is to clone, express, and determine the three-dimensional structures of many proteins by high throughput x-ray or NMR analyses with the added benefit of defining all folding motifs and, in some cases, function of individual proteins.

Proteomics may be defined broadly as the study of all proteins (and alternatively spliced varieties) expressed by a genome, including the concomitant isolation, identification, structural determination (with post-translational modifications), interaction with partners (other proteins, lipids, nucleic acids), expression, developmental time courses, affects on biological responses, and functional properties. Implicit in this definition is the utilization of high throughput methods that approach the actual concomitant accumulation of these data. In practice, a proteomic approach may focus on limited aspects (e.g. identification and interactions) of this broad menu of protein attributes or, unlike a genomic approach, may focus on a more limited proteome from cell fractionation such as that complement related to a cellular organelle (e.g. mitochondria) or component (e.g. nucleosome). Frequently, identification of the ground rules that govern function gets bypassed in the rush to accumulate high throughput protein structures or to identify interactomes; however, many investigators do emphasize the importance of a quantitative assessment of function. Implied in much of the high throughput approach is that discovery-based research will surpass hypothesis-driven experiments as a means of scientific progress, again a point for contention.

Proteomics has focused primarily on technology involving separation (two-dimensional gels, liquid chromatography, mass spectrometry), structure (x-ray and NMR), and interactions (mass spectrometry, yeast two-hybrid, immunoprecipitations, combinatorial methods) and a shift toward studying hundreds or thousands of proteins at a time. A truly erudite, comprehensive understanding of the functioning of a living cell also requires a quantitative, dynamic description of the stoichiometry, the kinetics of formation, the energetics of the formation, and the functional consequences of each protein complex in a cellular pathway. Classical biophysical analysis to determine size, shape, and solution properties of proteins has been around since the time of viscosity measurements and sedimentation analysis in the analytical centrifuge (3). Biophysical studies in the first half of the 20th century gave birth to our concept of proteins as defined globular or fibrous proteins of exact composition that were amenable to a more refined structural analysis at the atomic level. We are now,
Biophysics and Proteomics

perhaps, past puberty and into the adolescent years of growth, some years before a mature understanding of everything there is to know about the full complement of proteins in an organism. From a purely structural viewpoint, maturity is well advanced. The question that will be dealt with in this article will be, “What is the role of biophysical methods in the study of proteins in the proteomic era?” The issues will deal with the information content of physical biochemical methods, dynamic versus static structural data, and interactions in solution.

Biophysical approaches can provide data, complementary to the detailed molecular protein structures from crystallography and high resolution NMR, that reveal insights into how proteins behave in solution and how they interact dynamically with each other. Useful techniques include hydrodynamic methods (analytical ultracentrifugation, viscometry, etc.), thermodynamic methods (light scattering, microcalorimetry, surface plasma resonance), and spectroscopy (fluorescence, circular dichroism (CD),1 electron paramagnetism). In this article we will discuss the proteomic approach of these techniques with experimental examples.

CONFORMATIONAL CHANGES IN SOLUTION

Spectroscopy

Conformations at low resolution and changes in conformation with different conditions have, traditionally, been studied by spectroscopic methods that are highly sensitive and rapid. The sensitivity means that only small amounts of protein are needed, and dilute solutions can generally be used. The rapidity or instantaneous measurement of the signal allows kinetics of the processes under study to be measured and suggests that high throughput analyses could be developed, if desirable. These classical and neo-classical methods include ultraviolet light absorption difference spectroscopy, intrinsic fluorescence, extrinsic fluorescence with covalent or noncovalent probes, fluorescence resonance energy transfer in many variations, CD, and Fourier transform infra-red spectroscopy. These methods suffer, of course, compared with crystallography or NMR by having much less information content and, therefore, provide only limited details about the actual three-dimensional structure of the protein. Nonetheless, these methods have traditionally provided an entrée into the examination of solution properties of proteins with small amounts of material in non-destructive experiments. With judicious utilization of time and temperature variables, spectroscopy provides useful data on kinetics and thermodynamics of the conformational changes. The question is the role of these methods in modern proteomic approaches with robotics and high throughput assays.

The advantage that spectroscopic methods possess is the ability to vary conditions. The knowledge of the kinetics of a conformational change is important in interpreting the role of the transition in a biological process. Furthermore, knowledge of the $\Delta G$ of activation ($\Delta G^*$) from the kinetics often provides an understanding of the transition state and kinetic pathway for the conformational process. A measurement of $\Delta G$, $\Delta H$, and $\Delta S$ from temperature dependence of a conformational change establishes the energetic relationship between the forms and, therefore, the ability of external conditions to be able to shift between isomeric forms. From x-ray or NMR studies one may get two or three atomic resolution three-dimensional structures with different ligands bound; from solution spectroscopy one may get a continuous variation of distribution of structures between several conditions, e.g. ligand binding, pH, or temperature.

Stability

One particular variation of this procedure is the measurement of the stability of a native structure by measuring the $\Delta G$ for unfolding by temperature or denaturant, as monitored by spectroscopy (4). Would it be possible to automate the procedure for $\Delta G$ of stability determination via high throughput? Probably. Would it be useful to have the $\Delta G$ of stability determined for as many proteins as possible? Maybe. First, automation of a spectrophotometer/fluorometer that has been programmed to increase temperature or add a protein pre-incubated with denaturant would be quite feasible in the modern electronic/robotic era. The feasibility would be dependent on the cost and the demand for such large-scale measurements. Second, thermodynamic stabilities of more than a hundred different proteins have already been determined. General theories have been made based on these measurements (5), and more values would be helpful to extend such theories. However, because the general principles of folding and stability are fairly well understood (6), accumulation of more numbers may not provide more general insights. In this case, the understanding of individual proteins may benefit from $\Delta G$ stability measurements but should probably be done on a case-by-case basis or in a closely related family such as T7 lysozyme or gene V protein (7, 8) with a series of mutations to compare changes in stability.

A good example of a potential high throughput spectroscopic method to scan many proteins rapidly is the use of Congo Red (or thioflavine S) binding, and particularly the birefringence of binding, to determine the presence of amyloid structure in proteins (9). Especially relevant to this discussion is the use of Congo Red binding, along with fluorescence and CD, with those proteins that undergo a reversible transition between an amyloid form and a non-amyloid form, e.g. the prions (10, 11). Similarly, binding of 1-anilino-8-naphthalenesulfonate, or a similar fluorophore, to hydrophobic protein surfaces is often taken to indicate partially unfolded molten globule regions of proteins (12). Indeed, such methods could be automated to provide a high throughput survey of a family

---

1 The abbreviations used are: CD, circular dichroism; ECD, extracellular domain; NGF, nerve growth factor.
of proteins for amyloid- or molten globule-forming transitions subsequent to separation by high pressure liquid chromatography or other methods. A situation might be delineated in which such an analysis would be warranted but would represent a limited set of all proteins, e.g. pathological or neurological.

**MACROMOLECULAR ASSEMBLY**

Biological activity is the product of interactions between biological molecules. As a consequence of these interactions biological signals are transmitted through intra- and intermacromolecular interactions. At present there are very active proteomic research efforts to identify the pathways of signal transduction by identifying the biomacromolecular partners. Two of the more popular approaches are yeast two-hybrid assays (13–16) and mass spectrometry (17). However, identification of interacting partners alone is not enough to understand biology. How do these proteins interact in the complex? What is the order of interaction? What are the energetics that govern the interactions? What are the functional consequences of these interactions? What is the pathway of message transmission? Answers to these issues are essential for establishing the stepwise pathway of a biological function in the context of the cellular milieu in which these macromolecules reside. To address these issues the mechanisms of transmission of information and communication between macromolecules must be elucidated. These mechanisms include macromolecular assembly (protein-protein, protein-nucleic acid, protein membrane) and interfacial interaction (domain-domain). Furthermore, the pathways of propagation of energetic changes induced by ligand binding and macromolecular assembly need to be defined. Some of the approaches that can provide data pertinent to these issues are considered in the context of the proteomic era.

**Transport Techniques**

The size and shape of molecules can be estimated by monitoring the movement of these molecules in a force field such as centrifugal force for sedimentation, gravity for gel filtration, and electrical potential for electrophoresis. The principles that govern these techniques are the same, and thus the information content derived from these techniques is similar. The choice of a specific technique depends on the system that is under investigation, e.g. stability, availability of material, and equilibrium constant of the reaction. These approaches provide the information to monitor macromolecular assembly processes. These classical approaches are based on rigorous theoretical principles but have fallen from favor mainly because of the difficulties in using the instruments, the relative slowness of data acquisition, and the complexity of analyzing the data. However, with the recent advent of the new generation of instruments with state-of-the-art technology, these methods have become user-friendly. Furthermore, the availability of powerful software has enabled the users to extract the quantitative parameters with relative ease and speed. The reaction boundary contains data to provide information on the stoichiometry and equilibrium constant of macromolecular complex formation (18). This approach has been applied successfully to monitor the association between different transcription factors Smad 3 and Smad 4 (19), to determine assembly mechanisms of kinesin motor domains (20), to detect the presence of heterogeneity of assembly in β-lactoglobulin because of genetic variance (21), and to detect isomerization of conformational states in aspartyl transcarbamylase (22).

Gel filtration, particularly the large zone approach, provides the same information as the reaction boundary in sedimentation (23–25) but has an added advantage in its flexibility in methods of detection of the elution profile. Fluorescence, radioactivity, or enzyme activity may be used to monitor the elution of samples. Thus, gel filtration under the appropriate circumstances can extend the capability to monitor dissociation constants accurately to values as low as 10^{-15} M (26).

**Light Scattering**

Light scattering has undergone a period of renaissance as a result of new advances in instrumentation and development of new software. Current solution filters are very efficient in eliminating the chronic problem of dust particles. With the availability of a large number of interesting biological systems, this approach should again be considered seriously for studying macromolecular assembly, in particular, the kinetics of assembly processes (27, 28). A potential advantage of this approach is the distinct possibility of automation leading to a high throughput type of analysis. Dynamic or quasielastic light scattering technology with automated data analysis has improved in recent years (29) and provides a complement to static light scattering for determining radius of gyration and diffusion coefficient, as well as molecular mass (see below). Dynamic light scattering is currently used extensively to determine homogeneity and, hence, the capability of the protein to crystallize.

**Fluorescence Anisotropy**

Interaction between macromolecules can be detected by monitoring the changes in the steady state fluorescence anisotropy of a fluorescently labeled species. Basically, fluorescence anisotropy is a measure of the tumbling motion of the fluorescent probe and the labeled macromolecule. An increase in size of the labeled macromolecule through binding to another macromolecule can be measured directly by fluorescence anisotropy. The observed value of anisotropy is weighted in accordance to the fraction of free and bound labeled macromolecule. Thus, fluorescence anisotropy is a versatile technique to monitor macromolecular assembly and had been employed to monitor DNA-protein interaction (30–32), assembly of the DNA replication machinery (33), dynam-
ics of domain-domain interactions (34, 35), and protein-protein interaction (38). This approach has been developed recently as a high throughput assay for G protein-coupled receptor binding (37).

**Mass Spectrometry**

New developments in mass spectrometry are particularly useful in studying and quantifying the dynamics of macromolecular assembly. The approach of nanoflow electrospray coupled with time-of-flight mass spectrometry (38) allows one to study macromolecular assembly in its native state, and the range of molecular weights of the complex has been extended to $10^6$ Da, e.g. *Escherichia coli* ribosome (39). Not only can the mass of the final product of the assembly process be determined, but the approach has been applied successfully to define the macromolecular organization of the Yersinia pestis capsular F1 antigen (40), the chaperone complex from Methanobacterium thermoautotrophicum (41), the post-synaptic density-95 complex (42), and the yeast nuclear pore complex (43).

The versatility of the approach enables one to define the noncovalent pathways of assembly and disassembly of these macromolecular assemblies (44, 45). In the study of *E. coli* ribosomes, Robinson and co-workers (45) were able to detect the dissociation of some ribosomal proteins from the complex. Although at this stage of development thermodynamic parameters have not been derived from these observations, the mass spectrometric observation correlates qualitatively with the known reactions of these species (45). In other studies using electrospray ionization mass spectrometry, conditions have been worked out to determine thermodynamic parameters, i.e. the $K_d$ of the interaction. The $K_d$ values of the *E. coli* replication-inhibiting protein Tus and its mutants for the specific Ter DNA sequence in the nm to μM concentration range were determined (46) in agreement with classical solution studies. Thus, these developments are most encouraging to investigators who are interested in monitoring the pathways of signal transmission between macromolecules and the perturbation of the pathways as a consequence of further interaction with ligands or other biomacromolecules.

**Other Methods and Combinations of Approaches for Protein Characterization**

New technology, such as biosensors based on microfluidics and surface plasmon resonance (47), and revitalized methods, such as computerized isothermal titration microcalorimetry (48), also provide rigorous and powerful methods to determine dissociation constants and/or kinetics of macromolecular interactions. A recent paper (47) has documented that judicious use of the biosensor to measure binding on a surface can provide data equivalent to, or better than, solution methods of analysis. Robotics has maximized throughput of the Biacore biosensor and could potentially do the same for microcalorimetry.

A combination of these methods is frequently the most effective means to characterize proteins. Dynamic light scattering studies, along with gel filtration chromatography, analytical ultracentrifugation, and chemical cross-linking experiments, support the model that in *vivo* copper loading of yeast superoxide dismutase SOD1 occurs via a heterodimeric intermediate with its copper chaperone, rather than between two homodimers (49). Stopped-flow, static, and dynamic light scattering studies of vesicular stomatitis virus nucleocapsid proteins were used to determine their conformation, extent of self-association, and amount of bound matrix protein from the radius of gyration, concentration dependence of the apparent molecular mass, and diffusion coefficient (29). Dynamic light scattering, sedimentation, equilibrium, and circular dichroism measurements provided evidence that the anti-apoptotic protein BAG-1 exists as an elongated, highly helical monomer in solution (50). Isothermal titration microcalorimetry confirmed the 1 to 1 stoichiometry of the heterodimer with Hsp70, and both microcalorimetry and surface plasmon resonance yielded a $K_d$ of 100 nM for the complex that modulates chaperone activity (50).

**Receptors and Receptor Extracellular Domains**

Dimerization is viewed frequently as either a prerequisite or a required step in receptor activation, and receptors frequently bind several other proteins upon activation. The proteomic approach has been successful in identifying bound cellular proteins, e.g. 77 proteins have been identified as binding to the N-methyl-D-aspartate receptor complex (51). However, study of many individual receptor systems has focused on quantitative analysis of the association of the receptor extracellular domain (ECD), either with or without its ligand, that frequently provides the driving force for receptor dimerization. These latter studies demonstrate the types of quantitative information required for understanding how hormones and growth factors trigger cell signaling. Demonstration of an interaction or the determination of the three-dimensional structure by crystallography or NMR, although an extremely useful initial step, does not provide the entire picture without a more quantitative assessment.

Nerve growth factor receptor (NGF) and the neurotrophin family that includes NGF, brain-derived neurotrophic factor, neurotrophin 3, and neurotrophin 4 initiate cellular responses by binding to a series of cognate protein tyrosine kinase receptors TrkA, TrkB, and TrkC (52). Several studies have examined the interaction of the dimeric neurotrophin with the recombinant full-length ECD of the cognate receptor. Evidence has been provided from sedimentation equilibrium analysis and size exclusion chromatography that the full ECD of TrkB and TrkC form a complex with BDNF and NT3, respectively, with a stoichiometry of either 2:1 or 1:1 (receptor to neurotrophin dimer) depending on concentration (53). However, the dissociation constants determined were about three orders of magnitude higher than that expected from bioassay
and cellular binding studies. Also, size exclusion chromatography, in conjunction with light scattering, has indicated that the TrkA ECD is monomeric and dimerizes in the presence of NGF (54). This interaction coincided with a small but significant conformational change as determined by CD studies. Although binding of NGF to TrkA-ECD was measured with a surface plasmon resonance biosensor, no direct determination of the $K_d$ for receptor dimerization was measured in these latter studies.

In contrast, crystallographic studies have shown that the d5 IgG-like subdomain of TrkA forms an incorrect dimer in which symmetric domain swapping has occurred such that one $\beta$-strand of one TrkA chain has folded out to form a strand with the same $\beta$-sheet in the accompanying subunit (55). Such an interaction would not be able to bind NGF and would thus provide incorrect binding constants when used for solution studies. Whether such a false dimer will occur in the full-length ECD, which includes two cysteine subdomains, a leucine-rich subdomain, and two IgG-like subdomains, is unknown. Additional studies in solution of full-length ECDs of the Trk receptors should provide valuable information in this regard and correlate the structural and the biophysical data.

CONCLUSION

All biological functions are the resultants of complex linked reactions, frequently between biomacromolecules. As a consequence, a careful dissection of these linked reactions is needed before one can begin to define the ground rules that govern these reactions. Without a clear definition of these rules one can only describe qualitatively the biological phenomena. One needs to understand the quantitative aspects of these underlying principles before one can successfully engineer the biological functions to accomplish one's goals, such as discovering proteins involved in and then treating diseases. The quantitative parameters that define the basis for interactions and linked reactions are available only as a consequence of a meticulous course of study involving chemistry, biophysical properties, and structure. Such studies will often utilize focused, hypothesis-driven approaches. Nevertheless, biophysical characterization is challenged to keep up with the pace with which macromolecular structures, and gene expression data are being obtained. Thus, high throughput quantitative measurements of macromolecular assembly continue to be needed. Some of the approaches discussed above are amenable for further developments as high throughput assays; such efforts should be encouraged. The ultimate challenge is in the interpretation of the messages embedded in all the quantitative parameters describing the interactions of biomacromolecules.

Acknowledgment—Marc J. Glucksman is gratefully acknowledged for a critical reading of the manuscript.

* This work was supported in part by National Institutes of Health Grants NS24380 (to K. E. N.), NS36700 (to K. E. N.), and GM-45579 (to J. C. L.) and by Robert A. Welch Foundation Grants E-013 and E-1238 (to J. C. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Finch UHS/Chicago Medical School, 3333 Green Bay Rd., N. Chicago, IL 60064, E-mail: neetk@finchcms.edu.

REFERENCES


Biophysics and Proteomics


