Perspective

A Natural Interaction: Chemical Engineering and Molecular Biophysics

Abraham M. Lenhoff
Dept. of Chemical Engineering, University of Delaware, Newark, DE 19716

Introduction

Chemical engineers have carved several types of niches at the interface between fundamental biology and applications, occasionally venturing into direct competition with biologists on their turf, but more often extending the fundamental biology into areas with which they are better equipped to cope than life scientists. That chemical engineering is an increasingly molecularly oriented profession makes a molecular focus quite natural for much of life sciences research in chemical engineering as well. Indeed, there are parallels in some of the issues addressed by chemical engineers and biologists; for instance, the complex reaction pathways analyzed by chemical engineers are analogous to the metabolic pathways worked out by biologists during the first half of the last century. However, the driving forces for molecular-level studies in chemical engineering and the life sciences diverged substantially over the past half-century.

Chemical engineers, having the luxury of dealing with relatively simple molecules, have used the molecular level as the starting point for integrating over ensembles of molecules to determine continuum and ultimately macroscopic properties for design purposes. Biomolecular science, in contrast, undertook a reductionist quest to understand the structure and function of the molecular machinery of living systems. The result is an enormous amount of structural information on complex biomolecules and an extensive understanding of how they work. In addition, experimental approaches within various branches of what is now sometimes referred to as -omic science are producing prodigious sets of genetic information.

With such an extensive amount of structural and functional information on hand, the collective *Weltanschauung* of the biological community is now returning to a more integrationist one with a focus on how the components studied within the reductionist philosophy work together to produce a living organism. Although correlations within large experimental data sets may be helpful in solving these problems in the new biology, a *quantitative* analysis of biomolecular *systems* is ultimately essential, and there is no question that chemical engineers are uniquely well suited to this. This is seen in the modern practice of metabolic engineering (Bailey, 2001; Stephanopoulos, 2002), which is a natural offshoot of the biologists' mapping of metabolic pathways many years ago. Even this, however, does not cover the most impenetrable aspects of living systems, which are excellent examples of a wide range of complex physical and biological systems (Ottino, 2003). A chemical engineering background, in which the core chemical engineering science concepts can be effectively integrated with knowledge of the basic life sciences, is ideal for addressing a wide variety of such fundamental problems, as well as related applications.

Such efforts to elucidate the behavior of living systems, and to manipulate it via genetic approaches, have captured most of the attention in bioengineering over the past decade. Much less of the limelight has been enjoyed by a different interface between chemical engineering and biology, namely biophysics, which studies the physics and physical chemistry of biological molecules, molecular assemblies and cells. Molecular biophysics deals with *molecular structures*, especially of biological macromolecules such as proteins and nucleic acids and of assemblies such as lipid bilayers, and the relationships of these structures to *function*, both biological and physical. It is the latter in particular that has historically provided a natural synergistic relationship between molecular biophysics and chemical engineering, via chemical engineers' requirement for biophysical property data for bioprocess analysis and design, analogous to that for physical property data for designing conventional equipment to process simpler molecules. The long-standing relation between chemical engineering and molecular biophysics is probably due in part to the common origins of the disciplines in physical chemistry, as is well illustrated by early work on proteins. Indeed, one has to look at little other than the still-relevant volume of Cohn and Edsall (1943), an encyclopedic treatise on the state of protein molecular biophysics at the time, to appreciate how much the two disciplines have in common. The chapters deal with a wide range of theoretical and experimental topics that in modern chemical engineering would be covered in solution thermodynamics, transport phenomena and colloid science. A particularly interesting step that Cohn, Edsall, and colleagues took beyond their fundamental biophysical studies into the acknowledged domain of chemical engineering is that, as part of the war effort, their expertise was brought to bear on the problem of fractionating blood plasma via what became known as the Cohn process.

The timelessness of the concepts covered by Cohn and Edsall is remarkable in view of the very sparse information on protein structure then available. Although biophysics has since progressed to much more realistic representations of molecular structure, many of the principles enunciated earlier are still widely accepted, and are familiar topics to chemical engineers. There remain, however, substantial research challenges for both chemical engineers and biophysicists in integrating the principles and the structural information into methods for solving everyday problems. Much of the remainder of this article is devoted to exploring the reasons for this, using as a vehicle a paradigm of modern chemical engineering: estimation of physical properties from molecular structure, and use of these results for engineering design. This is routinely applied in traditional processing systems, for instance, use of molecular structure of hydrocarbons to determine vapor-liquid equilibria

that can be used to design a distillation column, while, in contrast, empiricism rather than a systematic approach such as this remains common in the design of bioprocess systems. This difference is due largely to a lack of suitable methods for estimating physical properties from molecular structural information. The reasons for this are explored in the remainder of this article with specific reference to proteins, where the problem of property prediction includes, in general, the problem of structure prediction, which is considered separately. For both of these, an initial brief overview of protein structure and properties is provided to set the stage.

The focus on property estimation and bioprocessing may provide a misleadingly narrow outlook on the future potential of molecular biophysics applications in chemical engineering. Molecules such as proteins can be harnessed as molecular machines of various kinds, with the more modern chemical engineering goal of molecular design as a component. These more ambitious applications are addressed in the final section.

Protein Structure and Properties

The distinctive properties of proteins that are of interest to chemical engineers result from their complex, but well-defined, structural features (Figure 1). A protein is a heteropolymer of amino acids, the unique character of which is determined by the amino acid sequence, or primary structure (Figure 1A), which is encoded in a gene (a segment of DNA) and, thus, represents "digital" information. Degrees of polymerization of tens to hundreds are typical, corresponding to molecular weights of order 10^3–10^5. The remaining features of the protein structure represent "analog" information in that the amino acid chain folds into a well-defined 3-D structure if the solution conditions are conducive to this. Although about 20,000 of these structures have been determined, mainly by x-ray crystallography, the question of how the folded state is reached (the protein folding problem) has been a major research area for numerous decades. The folded structure can be dissected into a number of local folding motifs, referred to collectively as secondary structure; the best known of these features are α-helices and β-sheets (Figure 1B). The overall 3-D form of a single polypeptide chain is the tertiary structure (Figure 1C), but sometimes multiple subunits self-assemble into a larger object such as a viral envelope (quaternary structure).

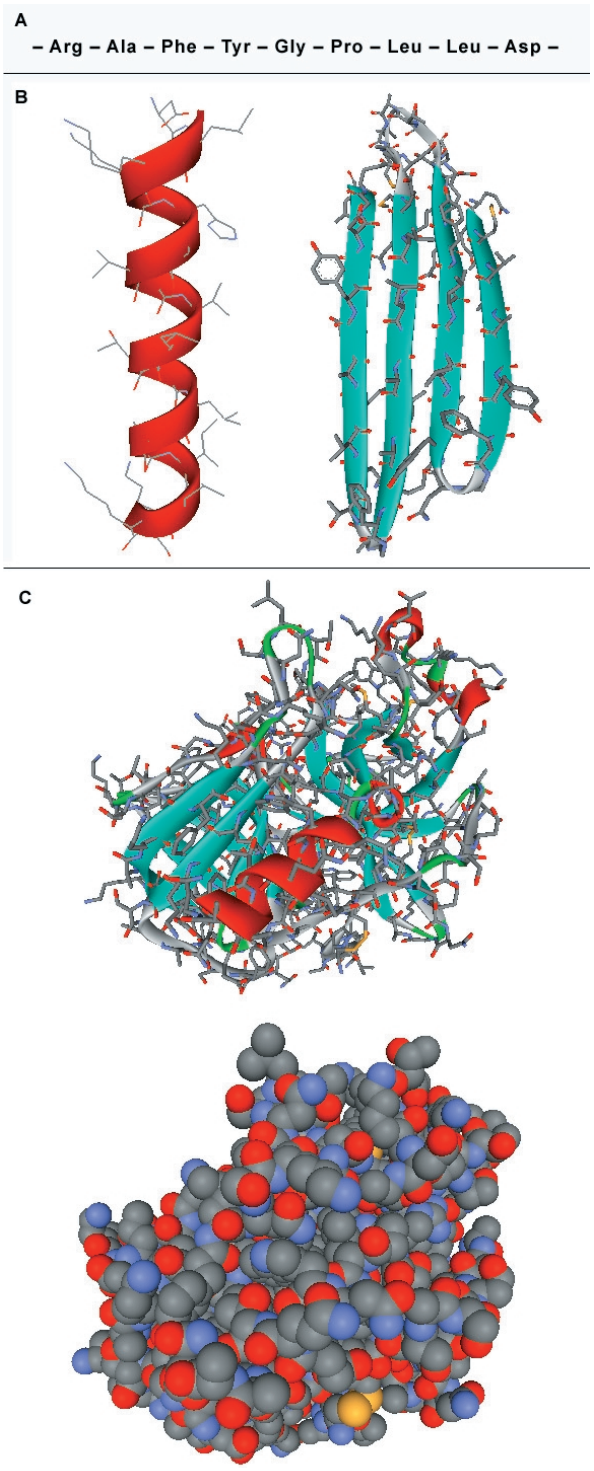

Figure 1. Organization of protein structure.
(A) Primary structure–the sequence of amino acids. (B) Secondary structure–local folding motifs such as α-helix (left) and β-sheet (right). (C) Tertiary structure–overall 3-D structure, shown as assembly of secondary structure motifs (upper) and in space-filling representation (lower) that illustrates the compactness of the folded structure. Atomic coordinates for protein structures are available from the Protein Data Bank at the Research Collaboratory for Structural Biology (http://www.rcsb.org/pdb/), where numerous downloadable software packages are also available for viewing the structures.

The physical properties of proteins are determined by the convolution of the properties of the individual amino acids and those of the overall folded structure. Several of these can be estimated by very straightforward back-of-the-envelope calculations that provide values useful in initial engineering design calculations:

• *Size and shape*: Folded globular proteins adopt remarkably compact structures with partial specific volumes that almost invariably lie in the range from 0.70–0.75 cm^3/g. Thus, given the fact that the amino acid sequence is usually known, the molecular weight, the molecular volume and the equivalent spherical diameter can be calculated successively. Protein effective diameters are on the order of a few nm to ca. 10 nm, which identify proteins as small colloidal particles, and key properties follow trivially. Diffusivities can be estimated quite well by the Stokes-Einstein relation, although these calculated values are typically slightly higher than the experimental values. This is because the molecules, although globular, are irregular rather than perfectly spherical, so their effective hydrodynamics diameters are larger than suggested by

equivalent spheres. Numerical calculations of hydrodynamics around an individual lysozyme molecule, as represented by the detailed crystal structure, yielded values of diffusivities within experimental error of the measured values (Brune and Kim, 1993), but the simpler estimation methods provide results that are well within the bounds of uncertainty with which engineers deal routinely.

• *Charge*: Another key colloidal property of proteins is that they are, in general, charged, by virtue of the ionizable character of a subset of the amino acid side chains. Some of these side chains are acidic and others basic, so that there is a strong pH dependence of the protein net charge. The net charge can be estimated simply and reasonably accurately from knowledge of the amino acid composition and the textbook pK values of the side chains. Such estimates of net charge are often useful in making qualitative predictions of protein behavior or heuristic design choices, e.g., in selecting adsorbents and pH operating conditions for ion-exchange separations.

These rather simple approaches to estimating some protein properties belie the appreciable challenges in estimating others, or even in predicting their qualitative behavior. The origins of most of these difficulties lie in the nature of intra- and intermolecular interactions in proteins (Elcock et al., 2001; Leckband and Israelachvili, 2001), for the calculation of which atomistic and colloidal models often represent the two extremes in a trade-off between realism and tractability, respectively. This conundrum can be seen by examining three key types of interactions:

• *Van der Waals interactions*: Van der Waals interactions, generally attractive, are present between all materials. At the atomic level, the Van der Waals interaction energy is usually represented as having an r^{-6} dependence on the separation distance r, familiar from the Lennard-Jones potential. This very short range results in a very strong dependence on the shape of the interacting surfaces that is not captured by the idealized shapes assumed for colloids (Roth et al., 1996).

• *Electrostatics*: The important role of electrostatic interactions in protein biophysical chemistry is reflected in the use of pH and salt concentration and type as the parameters most commonly used to manipulate protein solution behavior. Analysis of charge effects is complicated by the heterogeneity and anisotropy of the charge distribution, and by its dependence on pH. The presence of large numbers of free salt ions in the solution makes a direct simulation (primitive model) untenable. The compromise solution is a continuum description of the solvent in which the electric potential field is found as the solution to the Poisson-Boltzmann equation, in which the salt concentration is explicitly accounted for. This is analogous to what is used in colloid science, but the molecular and charge geometries present severe complications. However, the Poisson-Boltzmann approach has been both popular and successful for proteins in many cases for multimolecular systems and many monomolecular systems (Sharp and Honig, 1990).

• *Hydration interactions*: The unique physical properties of water are manifested in several respects in protein interactions, as enshrined in observed behavior such as the high dielectric constant of water and hydrophobic interactions. These effects can be effectively manipulated experimentally by the use of numerous additives, including salts, sugars, and denaturants such as urea. However, the phenomena are incompletely understood, primarily because they are a direct manifestation of the detailed molecular structure of water, and only explicit simulation approaches are really able to represent this adequately. This is an important area of enabling research in which chemical engineers are actively involved (Paulaitis and Pratt, 2002; Bystroff and Garde, 2003), but even the most advanced simulations have some way to go before efficiently capturing all aspects of behavior seen in real systems. Nonetheless, it is clear that hydration effects are, like van der Waals interactions, very sensitive to local geometry on a very short length scale (defined by the size of a water molecule), and to local molecular characteristics.

The challenges, then, in modeling protein behavior effectively lie in the complexity of the phenomena, especially hydration, and the strong dependence on structure that is closely coupled to environment. A particularly important limiting case of this is the very strong attraction that can develop when extended regions of two exposed surfaces fit together closely because they are geometrically complementary in their local topography, and, thus, able to facilitate extensive interactions over a very short range (Pauling and Delbrück, 1940). Indeed, it is this *complementarity*, modulated by the properties of the exposed surface groups, that drives the strong attractions that are generally referred to as biospecific interactions, or molecular recognition (Elcock et al., 2001; Leckband and Israelachvili, 2001), as well as protein folding. The ability to model interactions is further complicated by the fact that complementarity can be enabled or disrupted by relatively small changes in local structure, driven by changes in the structure of the protein itself, or by binding or removal of additional small molecules, including water or ions.

These aspects of protein solution behavior are the main reasons for the incomplete state of protein solution property prediction, and explain the organization of this article. The sensitive dependence on structure dictates a need for structure prediction, while the complexity of interaction patterns is what complicates property prediction. These are surveyed in the following two sections.

Protein Conformation and Structure Prediction

The protein folding problem is one that has occupied biophysicists since the 1950s, and its significance is reflected in its sometimes being referred to as the second half of the genetic code. Most of the motivation is purely scientific, but protein folding is also an essential step in producing proteins in the biotechnology industry. This, of course, occurs effectively in vivo, but proteins expressed at very high levels may appear as dense, incorrectly folded aggregates, called inclusion bodies, within cells. They must then be recovered by isolating the inclusion bodies, solubilizing and diluting them, and then refolding them from the unfolded form. Conformational considerations in the form of less extensive structural perturbations are also prominent in bioprocessing, e.g., adsorption, as well as in health care, via diseases such as Alzheimer's.

As in simulations on much simpler systems, two limiting protein folding formulations are that of mapping out a realistic folding pathway, and that of seeking only the final folded (presumably equilibrium) structure. In either case, the nature and magnitude of the problem are defined by Levinthal's paradox (Levinthal, 1969): for a (fairly small) protein containing about 100 amino acids, there are about 10^{50} possible configurations, so that sampling even at about 10^{-13} s per configuration yields an estimate of about 10^{30} years for the molecule to sample all conformations before settling into its native state. A direct Monte Carlo or similar approach to finding the equilibrium structure is correspondingly prohibitive. In stark contrast, proteins typically fold on time scales in the subsecond to minutes range.

Thus, proteins clearly do not reach their folded states by random conformational exploration. Instead, folding is often observed

experimentally to occur as a single-step concerted transition between the unfolded and folded (native) states. This reflects a transition between the dominance of configurational entropy, as in a random coil polymer, in the unfolded state, and that of the enthalpic and entropic (primarily hydration) contributions that sustain the folded state and render configurational entropy unimportant. However, the free energy of folding is quite small (–5 to –15 kcal/mol), so the key issue in modeling may be one of specificity rather than energetics (Lattman and Rose, 1993).

The folding transition was first modeled successfully using conceptual models of a kind widely familiar to chemical engineers, namely lattice models (Dill et al., 1995), in the limiting case with just two kinds of monomers included, viz. polar and hydrophobic, with very simple pairwise interaction potentials. These kinds of models indicate that folding occurs by local "nucleation" and subsequent propagation of secondary structural motifs, which have a much more limited number of degrees of freedom and therefore are able to assemble into the overall three-dimensional (3-D) structure. The final folded molecule has the bulk of the polar residues exposed to the solvent and most of the hydrophobic residues buried in the interior, consistent with the general picture observed experimentally. Although the protein folding community is now more interested in modeling and prediction of realistic protein structures, as discussed below, lattice models can still be valuable in exploring more complex folding situations. A good example is aggregation during protein expression (Smith and Hall, 2001), where simultaneous accounting for multimolecular behavior is not yet feasible using realistic models.

More realistic modeling of protein folding has been shown over the past few years to be possible, with folding of peptides and small proteins simulated directly by molecular dynamics (Duan and Kollman, 1998). Since single, relatively small molecules are involved, atomistic representations, including explicit water molecules, are feasible. In addition to the limitation on system size that is imposed by computational capabilities, the range of real times that can be covered for these systems is limited to ca. the microsecond range.

In view of these limitations, it is fortunate that explicit simulation of the folding process is not needed to find the final folded conformation of a protein. The development of alternative methods is a highly competitive area, largely to address the needs of the field of structural genomics, which seeks to determine the structures of proteins encoded by the genes sequenced as part of the Human Genome Project and related efforts (Burley and Bonanno, 2002). The performance of different methods is assessed competitively and objectively via the Critical Assessment of techniques for protein Structure Prediction (CASP) series (Lattman, 2001), in which different groups endeavor to predict the structures of proteins for which experimental crystal structures are about to become available. The methods used cover a range of approaches. A true biophysical one is to minimize the free energy of the system within the framework of a suitable set of interaction potentials. This is the kind of large-scale optimization problem frequently addressed by chemical engineers, and this reality has indeed been exploited in formulating the protein folding problem (Klepeis et al., 2002), although uncertainties remain in the accuracy of the potential functions. As noted above, the free energy of folding is relatively small, and specificity may be as important an issue as energetics (Lattman and Rose, 1993), but it is likely that inaccurate potentials remain one reason that biophysically based folding algorithms may fail.

Alternative approaches that are widely used for predicting the final folded structures of proteins are based on approaches relatively devoid of biophysical information. Instead, they exploit correlations of various kinds between the sequence of unknown structure and a variety of statistical data sets and known structures (Schonbrun et al., 2002). For instance, some amino acids have a stronger propensity than others to form α-helices, so an extended sequence of such amino acids would be more likely to appear in a helical region of the final folded protein. The database of known 3-D structures is used with "threading" algorithms in which sequence stretches identical or homologous (functionally similar amino acids) to others in the database are assigned similar secondary and tertiary structural features. Methods such as these currently enjoy greater attention than more biophysically based ones, and they represent one area in which chemical engineers play little part, but in the long run a combination of approaches may turn out to be optimal.

In the absence of reliable methods for structure prediction, experimental approaches for structure determination and manipulation are valuable, and for chemical engineers it is often paramount to do so in the process environment. Complete predictions of structural behavior in these systems are not possible, but in their absence even coarse-grained heuristics generated from experimental studies can represent useful information. Biophysical tools such as NMR, EPR and Raman spectroscopy are used to measure protein conformation in such systems as precipitates and adsorptive separations (McNay and Fernandez, 2001; Tobler et al., 2001). Structural manipulation, on the other hand, is accomplished by amino acid substitution using the methods of molecular biology or by a variety of thermodynamic variables. One that is commonplace for an engineer, but unusual for a biophysicist, is pressure, which affects both intra- and intermolecular protein interactions (Boonyaratanakornkit et al., 2002), again with the role of water being central (Hummer et al., 1998). Pressure has been shown to be useful in effecting recovery and refolding from inclusion bodies (Foguel et al., 1999).

Protein Interactions and Property Prediction

Folded proteins interact in vivo with surfaces such as cell membranes, with small molecules of different kinds (biological, organic, or inorganic) and with other biological macromolecules, including proteins, nucleic acids and carbohydrates. In the process environment, interactions with various kinds of surfaces may also be significant, and may be desired (e.g., adsorptive separations) or deleterious. Protein interactions with solvent, other protein molecules, and surfaces govern the physical properties on which bioprocess design is based. The significance and complexity of these interactions is sampled here mainly for protein-protein interactions as the determinant of solution thermodynamic properties and phase behavior.

The thermodynamic property most frequently examined for protein solutions is the osmotic second virial coefficient B_{22}, which is typically measured by osmometry or light scattering and for dilute solution is simply related to the activity coefficient (Grant, 2000). Osmotic properties are directly relevant in membrane separations, but B_{22} has also been shown to be correlated with the propensity of a protein to crystallize in a given solution (George and Wilson, 1994), thereby linking phase behavior with thermodynamic properties of the solution in the dilute limit. The formal linkage of interactions with bulk properties is seen in the statistical mechanical identification of B_{22} with the protein-protein potential of mean force, essentially the interaction free energy averaged over all solvent configurations (McMillan and Mayer, 1945). Although experimental data have been

fitted to relatively simple colloidal models (Vilker et al., 1981; Coen et al., 1995; Rosenbaum and Zukoski, 1996), these do not adequately capture the full range of the complexity of B_{22} behavior. What appears from several analyses to be a safe conclusion is that protein interactions and phase behavior are dominated by interactions of very short range (Rosenbaum and Zukoski, 1996; Lomakin et al., 1999), consistent with the discussion of interactions presented earlier. Implicating short-range interactions is significant because of the profound effect that the range of the dominant interaction has on the structure of the phase diagram (Anderson and Lekkerkerker, 2002). That a structurally and mechanistically realistic approach to predicting protein thermodynamic properties and phase behavior is still lacking, though, provides the framework for instructive insights regarding the complex and fascinating properties of protein solutions.

As discussed earlier, the strength of interactions is determined to a large extent by how closely the interacting moieties fit together over extended regions of their apposed surfaces. Interactions of proteins and other biomolecules are often classified as specific or nonspecific, but this is an artificial separation: the underlying biophysical mechanisms are the same, but strong and selective interactions, typically characterized by geometric complementarity, may be deemed to be specific. The profound dependence of protein interactions and their consequences on very small structural features is illustrated by the fact that sickle-cell anemia is caused by a single amino acid mutation in the 146-amino acid β subunit of hemoglobin. Strong protein-protein association is characterized by contact over extended regions, typically hundreds of $Å^2$ (Janin and Rodier, 1995), but these regions are not readily distinguishable from other parts of the protein surface because it is *pairwise* complementarity that drives the association. That the thermodynamic properties of such systems are dominated by such individual configurations is then a direct consequence of the Boltzmann weighting factor $\exp(-\Delta F/kT)$ that appears in calculations of solution properties by averaging over all configurations: the associations are characterized by free energies of association ΔF that can have absolute values of tens of kT (Horton and Lewis, 1992).

Such strong specific binding normally has a clear biological function, but it is also widely exploited in biotechnology, e.g., in the high selectivity of affinity separations. However, more generally in bioprocessing the protein interactions are not dominated by a single pairwise configuration, and the full distribution of free energies of interaction in different configurations becomes important. Calculations indicate that there will be a relatively small number of configurations of moderately high complementarity (Neal et al., 1998; Elcock and McCammon, 2001; Hloucha et al., 2001), and they contribute disproportionately to the Boltzmann-weighted average despite the fact that each is defined by a very small fraction of the overall 5-D angular configurational space that applies to pairwise interactions (Neal et al., 1998; Hloucha et al., 2001). These grossly anisotropic characteristics of protein interactions make spherically symmetric colloidal models unsuitable for calculating protein interactions, especially in including the colloidal representation of van der Waals interactions between spheres. Simpler model approaches to incorporating orientational dependence in protein interaction calculations have also been proposed (Lomakin et al., 1999; Sear, 1999), but they lack the pairwise complementarity that distinguishes protein-protein interactions from simply sticky patches. For protein interactions, a patch is sticky only when interacting with its distinctive "antipatch" (Hloucha et al., 2001), not with another generic sticky patch of the kind used in simpler models.

The argument that protein interactions are dominated by predominantly nonelectrostatic interactions within a small number of configurations appears to be in conflict with the well-established experimental observation that protein interactions may be dramatically altered by changes in pH and salt concentration and type, i.e., electrostatic parameters. This paradox, too, can be resolved with reference to the Boltzmann terms associated with the strongly attractive configurations (Neal et al., 1998). If the overall interaction free energy in such a configuration is made up of nonelectrostatic and electrostatic components ΔF_{ne} and ΔF_{es}, respectively, the Boltzmann factor can be written as $\exp(-\Delta F_{ne}/kT)\times\exp(-\Delta F_{es}/kT)$, with the first factor being a very large one. The electrostatic factor then serves as an amplification or attenuation factor, so even relatively weak electrostatic interactions can contribute quite large differentials in these configurations, and tuning of the electrostatics via pH or salt changes can lead to concomitantly large changes in the overall interaction energy.

Therefore, estimating the physical properties of protein solutions requires identifying pairwise configurations, in a 5-D configuration space, in which binding is strong, a challenge aggravated by the possibility of structural adaptations on binding (Davies and Cohen, 1996; Elcock et al., 2001). Docking algorithms to accomplish this (Sternberg et al., 1998) are now being assessed via the Critical Assessment of PRediction of Interactions (CAPRI) evaluation (Janin, 2002). Properties are then found by averaging over the full configuration space. The inherent complexity of the process is due to the strong anisotropy, the very short range of the interactions, and the very deep energy wells corresponding to the high-complementarity configurations. Each of these on its own represents a significant challenge to simulations, and, in combination, they present an inordinate level of difficulty. However, many of the obstacles present for simulations are related to behavior seen experimentally—slow kinetics, nonequilibrium phases such as gels and amorphous precipitates, multiple crystal forms, strong sensitivity to solvent properties—so these challenges to simulation offer a rich vein of both fascinating and informative science, as well as important practical benefits. Knowledge of the interactions is central to these, and in light of the difficulties inherent in direct prediction, the alternative avenue of efficient interaction measurements assumes greater importance (Tessier et al., 2002).

Prospects and Future Directions

The issues addressed above on estimation of structure and properties are essential ones for bringing analysis and design of bioprocesses onto the same footing as that of more conventional chemical processes. Nevertheless, quantifying and predicting protein folding and interactions remain largely on the horizon of current abilities, yet they do not do justice to the full range of potential implications and applications of protein biophysics, or to the intriguing fundamental scientific questions, that may be of interest to chemical engineers. These future directions must necessarily take us closer to relationships involving examples of the extraordinary array of more distinct biological functions possible. An indication of the versatility of proteins can be gleaned from a brief list of interactions occurring *in vivo*: enzymes (proteins that catalyze chemical reactions) bind to their substrates; protein binding to DNA regulates gene expression, with defects leading to cancer; the immune system functions via binding of antibodies to foreign species; oxygen is carried in red blood cells by binding to hemo-

globin; formation of ordered inorganic structures such as bone and teeth, as well as shells of marine organisms, is mediated by proteins. Ultimately, proteins play key and multifarious roles in the remarkable self-assembly and self-organizational properties of living systems. The mechanisms involved can often be dissected by integrating functional measurements with atomic resolution structures, but predicting function directly from sequence is not yet feasible in the absence of information from homologous systems. Direct design of specific functions, e.g., ability to transmit ions (Ghadiri et al., 1994), has been possible in some cases, but this is still most effectively done by mutating an existing molecule either by design or by directed evolution (Arnold, 2001).

Future progress in marshaling proteins in the service of humankind is likely to build further on analysis and synthesis of solution structure in general, and self-assembly in particular. Self-assembly of proteins can provide the foundation for a wide variety of structural and functional materials. Structural materials can include inorganic ones in which the template, as well as patterns of biomineralization, are directed ultimately by proteins, whether directly, or indirectly in conjunction with other structures such as vesicles (Mann, 2001), as is seen in the extraordinary array of structures found in diatoms (Figure 2). Functional materials, on the other hand, can cover the vast array of sophisticated and very specific functions that characterize protein function *in vivo*. An example of particular interest to chemical engineers is the broad class of membrane proteins, where self-assembly of what is often a multisubunit protein structure occurs within and adjacent to a lipid bilayer membrane. Alternatively, structural properties of proteins can be mapped onto functional arrays, in which the regularity of the structure, in 2-D (Nagayama, 1996) or 3-D, is exploited for its optical or electronic properties. Developing such structures systematically is a logical extension of an understanding of interactions involving proteins as well as other molecules, and of how to manipulate them, and one can foresee a range of levels of complexity of the structures formed by self-assembly. This will require the exquisite control of component concentrations, environment, rate and morphology that characterize the development of structure *in vivo*.

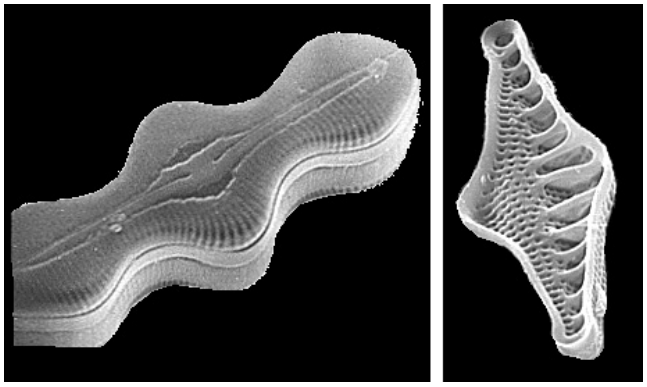

Figure 2. Diatoms as examples of inorganic structures formed by living systems.
Images shown (http://www.bgsu.edu/departments/biology/facilities/algae/html/Image_Archive.html) are used with the permission of Dr. Rex Lowe, Bowling Green State University.

A more significant challenge is that of understanding the structure of the intracellular milieu, the great complexity of which is reflected in the fact that even a simple bacterium such as E. coli has >4,000 different proteins, in addition to a vast variety of small organic molecules, lipids, and nucleic acids. Complex mammalian cells have significantly more proteins, as well as a greater degree of internal organization. The proteins and other components that make up the cytoskeleton, the walls of organelles, etc., serve to provide whole-cell characteristics that are manifested as mechanical properties of interest to cellular biophysics. Protein-protein interactions are also seen in enzyme complexes that facilitate rapid transport of metabolites among enzymes that catalyze successive reactions, a phenomenon known as channeling (e.g., Tsuji et al., 2001), or electron transport in complexes of mitochondrial proteins. The better defined and characterized of these kinds of complexes are, of course, firmly bound together by fairly strong specific interactions, but the possibility exists of proteins that tend to stay together in more nebulous complexes maintained by weaker interactions akin to those in simpler protein solutions. The importance of the resulting heterogeneity of the intracellular environment in maintaining the efficiency of metabolic systems remains to be assessed.

Extrapolating these arguments to their logical conclusion raises the enduring fundamental issues of the nature of living systems. As discussed earlier, chemical engineers have contributed much to organizing, in quantitative terms, existing knowledge of metabolic function, and both metabolic and genetic regulation are ripe for similar analysis. Such functional analyses are, however, restricted largely to *effect*, whereas *cause* resides in the symphony of molecules that make such effects possible, with proteins in the vanguard. Chemical engineers are well suited to undertake the large task of understanding and exploiting this as well, and to push the frontiers of knowledge to address the issue on which even the most blasé of scientists can reflect in awe: how, from a vast ensemble of molecules, each doing its own thing, can result in the miracle that is a living organism.

Acknowledgments

The insightful comments of Mike Paulaitis, Dilip Asthagiri, Julio Ottino and Pete Tessier during preparation of this article are gratefully acknowledged. Yong Duan and colleagues kindly prepared the images used on the cover, and Dr. Rex Lowe of Bowling Green State University generously consented to the use of images from his Web site in Figure 2. The research contributing to the preparation of this article was supported by NSF, NASA, and NIH.

Literature Cited

Anderson, V. J., and H. N. W. Lekkerkerker, "Insights into Phase Transition Kinetics from Colloid Science," *Nature*, **416**, 811 (2002).

Arnold, F. H., "Combinatorial and Computational Challenges for Biocatalyst Design," *Nature*, **409**, 253 (2001).

Bailey, J. E., "Complex Biology with No Parameters," *Nature Biotech.*, **19**, 503 (2001).

Boonyaratanakornkit, B. B., C. B. Park, and D. S. Clark, "Pressure Effects on Intra- and Intermolecular Interactions within Proteins," *Biochim. Biophys. Acta Prot. Struct. Mol. Enzymol.*, **1595**, 235 (2002).

Brune, D., and S. Kim, "Predicting Protein Diffusion Coefficients," *Proc. Natl. Acad. Sci. USA*, **90**, 3835 (1993).

Burley, S. K., and J. B. Bonanno, "Structuring the Universe of Proteins," *Ann. Rev. Genom. Hum. Genet.*, **3**, 243 (2002).

Bystroff, C., and S. Garde, "Helix Propensities of Short Peptides: Molecular Dynamics Versus Bioinformatics," *Proteins: Struct. Func. Gen.*, **50**, 552 (2003).

Coen, C. J., H. W. Blanch, and J. M. Prausnitz, "Salting Out of Aqueous Proteins: Phase Equilibria and Intermolecular Potentials," *AIChE J*., **41**, 996 (1995).

Cohn, E. J., and J. T. Edsall, *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, Reinhold Publishing Corp., New York (1943).

Davies, D. R., and G. H. Cohen, "Interactions of Protein Antigens with Antibodies," *Proc. Natl. Acad. Sci. USA*, **93**, 7 (1996).

Dill, K. A., S. Bromberg, K. Z. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan, "Principles of Protein Folding—A Perspective from Simple Exact Models," *Prot. Sci*., **4**, 561 (1995).

Duan, Y., and P. A. Kollman, "Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution," *Science*, **282**, 740 (1998).

Elcock, A. H., and J. A. McCammon, "Calculation of Weak Protein-Protein Interactions: the pH Dependence of the Second Virial Coefficient," *Biophys. J*., **80**, 613 (2001).

Elcock, A. H., D. Sept, and J. A. McCammon, "Computer Simulation of Protein-Protein Interactions," *J. Phys. Chem. B*, **105**, 1504 (2001).

Foguel, D., C. R. Robinson, P. C. de Sousa, J. L. Silva, and A. S. Robinson, "Hydrostatic Pressure Rescues Native Protein from Aggregates," *Biotechnol. Bioeng.*, **63**, 552 (1999).

George, A., and W. W. Wilson, "Predicting Protein Crystallization from a Dilute Solution Property," *Acta Cryst. D*, **50**, 361 (1994).

Ghadiri, M. R., J. R. Granja, and L. Buehler, "Artificial Transmembrane Ion Channels from Self-Assembling Peptide Nanotubes," *Nature*, **369**, 301 (1994).

Grant, M. L., "Effects of Thermodynamic Nonideality in Protein Crystal Growth," *J. Cryst. Growth*, **209**, 130 (2000).

Hloucha, M., J. F. M. Lodge, A. M. Lenhoff, and S. I. Sandler, "A Patch-Antipatch Representation of Specific Protein Interactions," *J. Cryst. Growth*, **232**, 195 (2001).

Horton, N., and M. Lewis, "Calculation of the Free Energy of Association for Protein Complexes," *Prot. Sci*., **1**, 169 (1992).

Hummer, G., S. Garde, A. E. García, M. E. Paulaitis, and L. R. Pratt, "The Pressure Dependence of Hydrophobic Interactions is Consistent with the Observed Pressure Denaturation of Proteins," *Proc. Nat. Acad. Sci.*, **95**, 1552 (1998).

Janin, J., "Welcome to CAPRI: A Critical Assessment of Predicted Interactions," *Proteins Struct. Func. Gen*., **47**, 257 (2002).

Janin, J., and F. Rodier, "Protein-Protein Interactions at Crystal Contacts," *Proteins Struct. Funct. Genet*., **23**, 580 (1995).

Klepeis, J. L., H. D. Schafroth, K. M. Westerberg, and C. A. Floudas, "Deterministic Global Optimization and ab initio Approaches for the Structure Prediction of Polypeptides, Dynamics of Protein Folding, and Protein-Protein Interactions," *Adv. Chem. Phys.*, **120**, 265 (2002).

Lattman, E. E., ed., Fourth Meeting on the Critical Assessment of Techniques for Protein Structure Prediction, *Proteins*, **45**, suppl. 5 (2001).

Lattman, E. E., and G. D. Rose, "Protein Folding—What's the Question," *Proc. Natl. Acad. Sci.*, **90**, 439 (1993).

Leckband, D., and J. Israelachvili, "Intermolecular Forces in Biology," *Quart. Rev. Biophys.*, **34**, 105 (2001).

Levinthal, C., in *Mössbauer Spectroscopy in Biological Systems*, P. Debrunner, J. C. M. Tsibris, and E. Münck, eds., University of Illinois Press, Urbana, IL (1969).

Lomakin, A., N. Asherie, and G. B. Benedek, "Aeolotopic Interactions of Globular Proteins," *Proc. Natl. Acad. Sci. USA*, **96**, 9465 (1999).

Mann, S., *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry*, Oxford University Press, Oxford (2001).

McMillan, W. G., and J. E. Mayer, "The Statistical Thermodynamics of Multicomponent Systems," *J. Chem. Phys.*, **13**, 276 (1945).

McNay, J. L. M., and E. J. Fernandez, "Protein Unfolding during Reversed-Phase Chromatography: I. Effect of Surface Properties and Duration of Adsorption," *Biotechnol. Bioeng.*, **76**, 224 (2001).

Nagayama, K., *Protein Array: An Alternative Biomolecular System*, Advances in Biophysics, Japan Scientific Societies Press, Tokyo (1996).

Neal, B. L., D. Asthagiri, and A. M. Lenhoff, "Molecular Origins of Osmotic Second Virial Coefficients of Proteins," *Biophys. J*., **75**, 2469 (1998).

Ottino, J. M., "Complex Systems," *AIChE J.*, **49**, 292 (2003).

Paulaitis, M. E., and L. R. Pratt, "Hydration Theory for Molecular Biophysics," *Adv. Protein Chemistry*, **62**, 283 (2002).

Pauling, L., and M. Delbrück, "The Nature of Intermolecular Forces Operative in Biological Processes," *Science*, **92**, 77 (1940).

Rosenbaum, D. F., and C. F. Zukoski, "Protein Interactions and Crystallization," *J. Cryst. Growth*, **169**, 752 (1996).

Roth, C. M., B. L. Neal, and A. M. Lenhoff, "Van der Waals Interactions Involving Proteins," *Biophys. J*., **70**, 977 (1996).

Schonbrun, J., W. J. Wedemeyer, and D. Baker, "Protein Structure Prediction in 2002," *Curr. Opin. Struct. Biol.*, **12**, 348 (2002).

Sear, R. P., "Phase Behavior of a Simple Model of Globular Proteins," *J. Chem. Phys.*, **111**, 4800 (1999).

Sharp, K. A., and B. Honig, "Electrostatic Interactions in Macromolecules: Theory and Applications," *Ann. Rev. Biophys. Biophys. Chem.*, **19**, 301 (1990).

Smith, A. V., and C. K. Hall, "Protein Refolding Versus Aggregation: Computer Simulations on an Intermediate-Resolution Protein Model," *J. Mol. Biol.*, **312**, 187 (2001).

Stephanopoulos, G., "Metabolic Engineering: Perspective of a Chemical Engineer," *AIChE J*., **48**, 920 (2002).

Sternberg, M. J. E., H. A. Gabb, and R. M. Jackson, "Predictive Docking of Protein-Protein and Protein-DNA Complexes," *Curr. Opin. Struct. Biol.*, **8**, 250 (1998).

Tessier, P. M., A. M. Lenhoff, and S. I. Sandler, "Rapid Measurement of Protein Osmotic Second Virial Coefficients by Self-Interaction Chromatography," *Biophys. J*., **82**, 1620 (2002).

Tobler, S. A., N. E. Sherman, and E. J. Fernandez, "Tracking Lysozyme Unfolding during Salt-Induced Precipitation with Hydrogen Exchange and Mass Spectrometry," *Biotechnol. Bioeng*., **71**, 194 (2001).

Tsuji, S. Y., N. Wu, and C. Khosla, "Intermodular Communication in Polyketide Syntheses: Comparing the Role of Protein-Protein Interactions to Those in Other Multidomain Proteins," *Biochemistry*, **40**, 2317 (2001).

Vilker, V. L., C. K. Colton, and K. A. Smith, "The Osmotic Pressure of Concentrated Protein Solutions: Effect of Concentration and pH in Saline Solutions of Bovine Serum Albumin," *J. Coll. Interf. Sci.*, **79**, 548 (1981).