FEATURE ARTICLE

Computer Simulation of Protein−Protein Interactions

Adrian H. Elcock,*,† David Sept,‡,§ and J. Andrew McCammon‡

Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242-1109, Howard Hughes Medical Institute, Department of Chemistry and Biochemistry, Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365

Received: October 2, 2000

The use of computer simulations in investigations of protein−protein interactions is discussed. First, crystallographic analyses of known protein−protein complexes are summarized with particular emphasis being placed on the atomic nature of the interactions. Models available for describing macromolecular association energetics are then discussed, with special reference to the treatment of electrostatic and nonpolar interactions. The use of these models in combination with efficient search methods is discussed in the context of the so-called protein docking problem and in the description of weaker (i.e., noncrystallizable) protein−protein interactions. Finally, simulations of the dynamics of protein−protein association events are outlined. In all cases, differences are stressed between the atomically detailed view of protein−protein interactions and the view implicit in the use of simpler colloidal models.

Introduction

Protein−protein interactions underpin many biological and biophysical processes. Examples from biology include the interactions involved in converting extracellular signals into appropriate cellular responses (signal transduction[1]), the control of gene expression by interactions between transcription factors,[2] the inhibition of enzymes by intracellular inhibitor proteins,[3] and the maintenance and regulation of the cytoskeleton.[4] More biophysical manifestations of protein−protein interactions are reflected in processes related to phase equilibria such as crystallization,[5] and in the marked dependence of the diffusion coefficients of macromolecules on their concentration.[6] The latter aspect is currently becoming of interest to more biologically minded researchers with the advent of techniques for monitoring protein diffusion in the crowded cellular environment.[7−9]

The purpose of this article is to outline the uses of computer simulations for describing protein−protein interactions. Although simulation techniques are many and varied, they can be broadly divided into two areas: (1) those that describe the proteins involved in atomic or near-atomic detail, with the aim of investigating the roles played by specific amino acids in the interactions, (2) those that treat proteins as simple geometric shapes (generally spheres), interacting by potential functions such as the DLVO potential,[10] derived for treating colloidal suspensions. Studies of the latter type are typically aimed at describing the effects of macromolecular interactions on the meso- or macroscopic properties of protein solutions, such as their diffusional behavior (see, e.g., ref 11). Although these studies are often of great interest, we restrict our attention here to studies of the first type. We make this choice for two reasons. First, the total breadth spanned by the two fields combined is so large as to defy adequate coverage in the allotted space. Second, even a cursory reading of the extant literature reveals that the distance between the two fields in terms of language is often embarrassingly large: an exchange of papers between workers from the two areas can in many cases be safely accomplished without any meaningful transfer of information. Because of this, one of our main hopes in writing this work is to provide a description of the methods used in more atomically detailed studies, that is sufficiently accessible that it will serve as a useful primer both for theoreticians more familiar with DLVO type models and for experimentalists interested in understanding protein−protein simulation techniques.

We begin by briefly reviewing the structural basis of protein−protein interactions as deduced from crystallographic studies of protein complexes. It will be obvious from this discussion that the view of protein−protein interactions that emerges is very different from that implicit in colloidal studies, since we place particular emphasis on specific types of atomic interactions. This emphasis carries over into the next section, where we discuss the kinds of energetic models that have been developed to describe the association energetics of protein−protein complexes. The use of these models in combination with efficient search methods is then discussed with reference to the so-called protein docking problem, which aims to predict the structures of protein−protein complexes. Next we discuss the use of similar methods to describe weaker interactions between proteins, which may not be observable by crystallographic techniques but, nevertheless, are manifested in another experimentally measurable quantity, the second virial coefficient.

§ Current address: Department of Biomedical Engineering, Washington University, St. Louis, MO 63130-4899.
† University of Iowa.
‡ University of California at San Diego.

10.1021/jp003602d CCC: $20.00 © 2001 American Chemical Society
Published on Web 02/01/2001

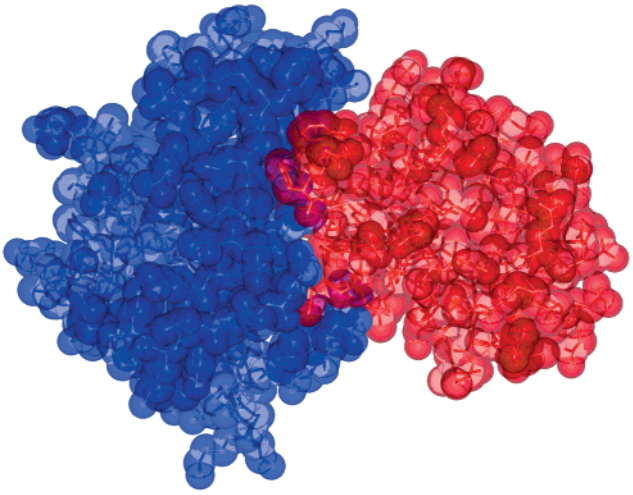


Figure 1. Surface complementarity between the enzyme barnase (blue) and its intracellular inhibitor barstar (red). For a fuller description of this structure, see ref 3. This figure was prepared using InsightII (Molecular Simulations, Inc.).

Finally, we outline the use of atomically detailed models in the simulation of the dynamics of protein−protein association events.

Analyses of Protein−Protein Interfaces

To develop an atomically detailed understanding of protein−protein interactions, we take as our starting point the analysis of the known crystal structures of protein−protein complexes.[12−17] Structural analyses are straightforward to perform, given a suitably representative sample of crystal structures, and are crucial for formulating ideas regarding factors likely to be important in energetic models for describing general protein−protein interactions. Perhaps not surprisingly, the types of interactions deemed to be important for complexing two proteins are much the same as those identified as critical for protein folding: the two processes are after all closely related (although not identical: see ref 18). Statistical analyses show, however, that there are differences between the relative importance of the various types of interactions, differences that are also apparent when one compares different types of protein−protein interactions.[13,19,20] Again, this should not be surprising: the evolutionary forces acting on the monomer−monomer interfaces of proteins that only exist as oligomers, for example, are quite different from those acting on the interfaces formed in complexes (e.g., antibody−antigen complexes[21]) whose component proteins spend much of their existence in isolation from one another.[20] For the following, we base our discussion on interactions of the latter type, which have been the subject of a very recent and thorough review.[17]

The first observation that can be made about protein−protein complexes may seem obvious: the surfaces of proteins often show a high degree of structural complementarity (Figure 1), with protrusions on the surface of one protein being matched with recesses in the surface of the second protein.[22,23] In the recent study of LoConte et al.,[17] the average area of protein−protein interfaces from a set of 75 crystal structures was ∼2000 Å^2, with an average of ∼200 atoms on each partner being involved in the interface. This structural complementarity is in many cases such a compelling hallmark of a specific protein−protein interaction, that a search for high complementarity forms the basis of most computer algorithms that seek to predict the structures of protein−protein complexes (see below). Not surprisingly, however, there is more to intimate protein−protein association than simply shape matching: the nature of the component amino acids of the interface is also critically important. On average, greater than half of the interface area is formed by amino acids with hydrophobic side chains:[17] these side chains, which are exposed in the unbound state, become juxtaposed with other hydrophobic side chains[24] and protected from water in the bound state (Figure 2A). Although the exact origins of the hydrophobic effect remain in dispute, the magnitude of the effect, and its role in promoting macromolecular association is well established (e.g., ref 25). In addition to the association of hydrophobic elements, other interactions that are clearly of importance to protein−protein interfaces include the formation of hydrogen bonding interactions (Figure 2B) between polar side-chain and backbone atoms (an average of 10 ± 5 in each complex[17]) and ionic interactions or salt bridges (Figure 2C) between oppositely charged amino acid side chains. The presence of the latter is the main reason why in addition to structural complementarity, there is also in many cases a pronounced degree of electrostatic complementarity between two bound proteins: regions of positive electrostatic potential on one protein match with regions of negative electrostatic potential on the partner.[26,27]

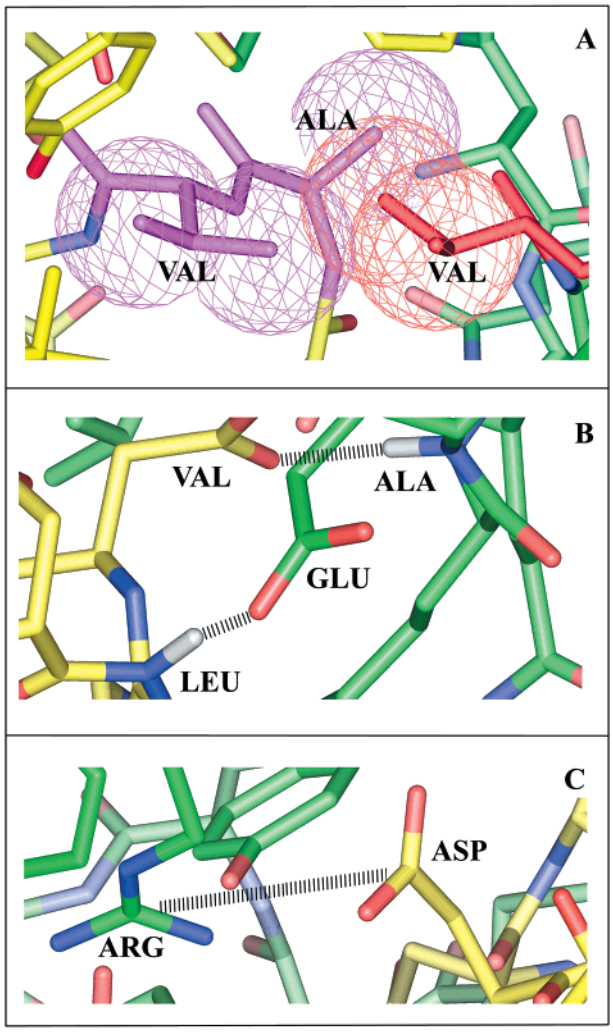


Figure 2. Examples of atomic interactions observed in protein−protein complexes: (A) close clustering of hydrophobic side chains; (B) hydrogen bonding interactions; (C) salt bridge (ion pair) interactions.

Lest the above discussion give the impression that protein−protein interfaces are so intimate that there is no free space left,

it is worth noting that water molecules are almost always identifiable in macromolecular interfaces[28] and in some cases, the interfaces can appear almost soaked in water. A prime example of this is seen in a protein−DNA crystal structure: that of the trp repressor protein complexed with its specific DNA binding site.[29] In fact, the complex is so riddled with water molecules in the interface, that some workers questioned the authenticity of the structure,[30] proposing instead that what had been isolated was a looser complex of the type that would have been observed between the repressor and *any* stretch of DNA (i.e., nonspecific binding). Following subsequent work, it has become generally accepted that the crystal structure is what it was originally claimed to be[31,32] and that specific protein−DNA (and by extension protein−protein) interactions can in fact be mediated by water.

In addition to the observation of direct and indirect interactions there is important information that can be gleaned from what is *not* observed in the crystal structures of protein−protein complexes. It is for example unusual to find charged amino acids buried in the protein−protein interface, which do *not* form either salt bridging interactions or extensive hydrogen bonding interactions with amino acids on the partner protein. In physicochemical terms, the reason for this is straightforward to understand. Because of the strongly favorable interaction of water with charged groups, a charged amino acid located on the surface of a protein will have a very negative (i.e., favorable) hydration free energy. Binding to a second partner protein that causes this charged amino acid to be transferred to the interior of the protein−protein interface, where it is removed from contact with the solvent, will therefore be an energetically unfavorable process and, if not compensated by the formation of new favorable interactions with the partner protein, would have deleterious consequences for the stability of the complex. Several examples have been reported in which charged amino acids incorporated into an otherwise hydrophobic protein interior have been shown to destabilize proteins.[33,34] This is therefore at least part of the reason for the above-mentioned electrostatic complementarity seen in protein−protein interfaces. Because unfavorable desolvation effects apply not only to charged side chains but also to neutral, polar groups in proteins, it is often argued that their interactions do not contribute strongly to the stability of isolated proteins;[35] this argument may not, however, apply so strictly to the formation of complexes with other proteins.[18] Even if the contribution of hydrogen bonding interactions is in dispute, it is generally accepted that they provide specificity to complexes, by providing a means of discriminating against alternative binding orientations that lack appropriate complementarity.[35,36]

The roles played by individual amino acid side chains in determining the strength of a protein−protein complex can be investigated experimentally simply by mutating the amino acid of interest (using the technique of site-directed mutagenesis) and measuring, through traditional calorimetric methods, the effect on the binding affinity. Particularly important has been the method of alanine scanning[37] in which the amino acids are mutated, one by one, to alanine, the side chain of which consists only of a single methyl group. Comparison of the binding affinity of the real (wild-type) protein with the alanine mutant protein gives an indication of the importance of that amino acid's side-chain's interactions for the binding affinity—under the assumption of course that removal of the side chain does not cause any drastic structural rearrangement of the complex. The idea of mutating amino acids to glycine, which has no side chain at all, although at first sight attractive, is in practice only rarely done because glycine has unique conformational properties (see, e.g., ref 38) that set it apart from all other amino acids and compromise facile interpretation of the results. A recent analysis of a series of alanine scanning results has indicated that so-called "hot spots" (i.e., amino acids whose replacement by alanine is especially unfavorable) appear in the centers of protein−protein interfaces where they are protected from solvent.[39]

In addition to these traditional thermodynamic approaches to understanding protein−protein complex stability, increasing work is focusing on exploiting the wealth of information that is becoming available from genome sequencing projects. For any given protein for which there is a three-dimensional structure in the database (http://www.rcsb.org), there are also many more evolutionarily related (i.e., homologous) proteins for which only the amino acid sequences are known. In principle, the existence of these sequences (which typically are from identical proteins isolated from different organisms) can provide information complementary to that obtained from conventional alanine scanning experiments. If we make the reasonable assumption that homologous proteins all bind to their protein partners in a more or less identical way, then comparison of their sequences can reveal which amino acids can be replaced freely without completely abolishing binding, and which amino acids must be maintained (conserved) in order to retain binding. In line with this hypothesis, studies have shown that residues involved in binding are clearly subject to increased degrees of conservation.[40,41] Further information can in principle be obtained from analyzing the correlated mutation patterns of amino acids *across* protein−protein interfaces.[42] Perhaps surprisingly, no truly in-depth study of such patterns has yet been published, although it seems unlikely that this will remain the case for much longer.

Calculating the Energetics of Protein−Protein Interactions

The prediction of the free energy of association (i.e., the binding affinity) of two proteins is one of the primary goals in the study of protein−protein interactions. For the moment, the most complete energetic descriptions of protein−protein interactions come from molecular mechanics (MM) methods: quantum mechanical (QM) treatments of entire proteins have started to appear recently but are not currently useable in a routine way.[43,44] MM formulations include terms representing the covalent and noncovalent interactions of all atoms of a molecular system; a number of excellent reviews of such techniques are available.[45−47] Several groups have expended considerable effort on parametrizing these methods specifically for the purpose of describing biomolecules immersed in aqueous solution (e.g., refs 48 and 49). When combined with molecular dynamics (MD) simulation methods (e.g., ref 45), these models provide a means of describing the often highly complex dynamical behavior of macromolecules in aqueous solution. With current computational resources, the time scale of such simulations is limited to tens of nanoseconds, although at least one simulation of 1 μs duration for a small protein fragment has been reported.[50] Such a time scale is too short to routinely observe processes such as protein folding,[51] but it is still often sufficient to observe smaller scale conformational fluctuations that are of relevance to a macromolecule's function, e.g. the gating motion of amino acid side chains involved in controlling access of substrate molecules to the active site of the enzyme acetylcholinesterase.[52] Because of this time scale restriction few MD simulations of protein−protein complexes have been reported, and although they may provide interesting information related to issues such as how

interactions between widely separated functional sites are mediated (e.g., refs 53 and 54), many aspects of the results must be considered rather speculative.

The easiest way to decrease the computational workload involved in assessing protein−protein interaction energetics is simply to carry out calculations on only a single conformation of the protein−protein complex, usually the structure found in the crystal. Since the solvent structure around macromolecules cannot be adequately described by a single snapshot (water is a liquid), the single conformation approximation is usually combined with a solvent model that in some way provides a description averaged over all the possible solvent configurations. For small solutes, integral equations can be solved to obtain the average water structure around the solute molecule(s), but for macromolecules such methods are still in development (e.g., ref 55). A more common approach is therefore to replace the explicit solvent molecule treatment with an implicit, continuum model (for a recent review of these methods, see ref 56). Generally, these models combine two terms: (1) a term that describes electrostatic contributions to solvation, usually through either a generalized Born[57] or Poisson−Boltzmann-based approach,[58] and (2) a term that describes nonpolar contributions to solvation, such as van der Waals interactions with the solvent, the energy required to form a cavity in the solvent large enough to accommodate the solute, and the hydrophobic effects associated with solvation of nonpolar solutes.

Electrostatic Interactions

The most popular approach to calculating electrostatic interactions in a solvated, atomically detailed, biomolecular system is to solve the Poisson−Boltzmann equation.[58]

$$\nabla\cdot[\epsilon(r)\,\nabla\Phi(r)] - \kappa^2(r)\sinh\Phi(r) = -4\pi\rho^f(r) \qquad (1)$$

where ∇ is the differential operator, Φ is the electrostatic potential, ϵ the dielectric constant, κ the Debye−Hückel parameter (related to the inverse square root of the ionic strength), and ρ^f is the charge density. *All of these terms are dependent on position*. This aspect is particularly important to remember for the dielectric constant, since the dielectric properties of proteins are very different from those of water. Controversy continues to rage about both the magnitude and the meaning of the term dielectric constant as applied to proteins, but there is at least agreement that it is the inhomogeneous structure of proteins that is the cause of difficulties in interpretation.[59−61] In the protein interior, which is generally occupied by nonpolar hydrophobic amino acids, the dielectric response (i.e., the response to an external electric field) is typically only small (∼2−4), being due more or less exclusively to electronic polarizability. At the protein surface on the other hand, there are many charged amino acids whose side chains are highly flexible. Since these side chains can reorient rather easily, their ability to respond to applied electric fields is much greater; a variety of theoretical treatments[62−65] have estimated their dielectric response to be in the range ∼10−30. Because of these differences, there is clearly no such thing as a dielectric "constant" for a protein and any value that is so defined must therefore be viewed as an empirical quantity that provides some kind of average description of the overall (more complicated) dielectric behavior. To emphasize its empirical nature, we can point to the fact that the optimum value to use appears to depend on the particular application: for calculating pK_a values of ionizable amino acids in proteins, for example, a value of 20 has proven highly successful,[66] whereas in other applications values of 1−8 seem more appropriate.[67] Attempts have been made to account for the varying behaviors by assigning different dielectric values to different protein regions (see, e.g., ref 68).

The other terms in eq 1 are easier to define. The charge density ρ^f represents the spatial distribution of the partial atomic charges in the macromolecule(s). Generally, these charges are taken from conventional molecular mechanics force fields such as AMBER[48] or CHARMM,[49] although attempts have been made to derive charges specifically for use in PB-based calculations (e.g., ref 69). The sinh term of eq 1 (often written in other derivations as a sum of exponential terms) is due to Debye and Hückel; it provides a mean-field description of the distribution of dissolved ions such as Na^+ and Cl^- in the solution and so allows the PB equation to be applied to solutions containing added salts. Implicit in the approach, then, is the assumption that while the atoms of the protein adopt fixed positions in space, the positions of dissolved ions (in addition to the water molecules) are thermally averaged and so can be represented by continuous distributions rather than by discrete locations. For small values of the potential, the sinh term can be Taylor-expanded and truncated at the first term; this approximation, which is generally only valid for low overall charge densities, yields the linearized PB equation.

Because the shapes of proteins are complicated, the Poisson−Boltzmann equation is usually solved numerically. Although boundary element methods have been used,[70,71] the most widespread methods use finite difference solvers[72,73] as implemented in the programs DelPhi[74] and UHBD:[75,76] the macromolecule(s) of interest is (are) placed at the center of a 3-dimensional (cubic) grid, and the equation is iteratively solved until satisfied to a given tolerance at all grid points. Since such calculations can take several minutes to perform for a macromolecule, efforts to accelerate the solution of the equation continue: a very recent development has been the introduction of fast adaptive mesh methods.[77,78] Whatever method is used, the result of solving the equation is to obtain the electrostatic potential throughout space, both in and around the macromolecule(s). By itself, this information can yield important insight into the function of a protein: as we shall see in a subsequent section for example, some enzymes generate substantial electrostatic potentials that are used to accelerate their association with other molecules. Because of this, programs such as DelPhi and UHBD are becoming standard tools for structural biologists wishing to examine the electrostatic properties of a newly solved macromolecular structure.

For the present purpose, a more important point is that once the electrostatic potential has been obtained it becomes straightforward to calculate the electrostatic free energy, ΔG_{elec}, of the macromolecular system (see Sharp and Honig's work[79] for expressions covering the calculation of ΔG_{elec} with the nonlinear equation):

$$\Delta G_{elec} = \frac{1}{2}\sum q_i\Phi_i \qquad (2)$$

where the summation is over all the charges in the macromolecule(s) and Φ_i indicates the value of the electrostatic potential at the position occupied by the ith charge. It is the ease of obtaining ΔG_{elec} that makes the PB equation attractive for understanding the association thermodynamics of macromolecular complexes. Calculation of the electrostatic contribution to protein−protein association can be made rigorously by performing three separate PB calculations: one of the complex, and one each of the two partner proteins. The electrostatic contribution to the binding free energy is then given by (ΔG_{elec}

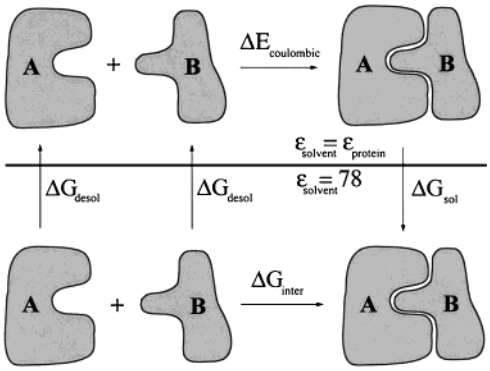


Figure 3. Thermodynamic cycle used for calculating electrostatic contributions to binding of two macromolecules A and B. See text for details.

of the complex) − (ΔG_{elec} of protein 1 + ΔG_{elec} of protein 2). It should be stressed that within the approximation of a single macromolecular conformation, fixed in a given protonation state, calculations of this type yield a *complete* description of the electrostatic contributions to binding. In other words, the screened Coulombic interaction between the macromolecules, and the effects of electrostatic desolvation (see below) are both implicitly included in such calculations. In most published applications, however, the PB equation is used only for calculation of the solvation components, and a thermodynamic cycle similar to that shown in Figure 3 is employed so that Coulomb's law can be used to calculate the direct interaction component. The calculation is accomplished as follows. The aqueous phase association process in which we are interested is conceptually broken into several stages, the first stage of which is the desolvation of the two (separate) molecules that results from transferring them from the aqueous phase to a medium that has the same dielectric value as the protein interior. The reason for transferring to such a state is that now, since the dielectric is spatially invariant, Coulomb's law can be used to analytically calculate the electrostatic energy change that results when the two molecules are brought together (ΔE_{coul}). Finally, the complex is resolvated by transferring it back to the aqueous solution. The solvation and desolvation energy components are calculated straightforwardly again using eq 1, as the difference between two calculations: one with the molecule immersed in an external dielectric of 2−20 (i.e., whatever value is assigned to the protein dielectric), and a second with the molecule immersed in an external dielectric of 78 (i.e., the aqueous phase).

The thermodynamic cycle approach is worth noting because it provides a straightforward illustration of the fact that the relative solvation energies of the complex and the isolated molecules can be important determinants of the overall binding affinity. We stress this because there is still a tendency in the literature to view binding electrostatics solely in Coulombic terms (i.e., if the interacting groups have opposite charges, then the interaction is often assumed to be favorable). Although this is probably generally true, it should be remembered that the approach of two proteins causes the high dielectric solvent around them to be displaced by a lower dielectric. This unavoidable "electrostatic desolvation" is the cause of the low likelihood, discussed earlier, of finding unpaired charged groups buried in proteins or interfaces. Unfortunately, estimating the magnitude of desolvation effects depends critically on the choice of the dielectric constant assigned to the protein:[80] the lower its value, the greater the desolvation effect. Given the uncertainties outlined above concerning what is the correct value to use, it is perhaps not surprising to find that very few calculations of the absolute contribution of electrostatics to protein−protein binding have been reported (e.g., ref 81). Instead, what is usually studied is the *change* in binding electrostatics that results from some change to the system (such as changing the ionic strength): the results of such calculations are typically much less sensitive to the value assigned to the protein dielectric constant. Good examples have been the calculations aimed at understanding the salt dependence of the stability of ligand−DNA complexes.[82−84] Other examples are provided by the use of PB methods to assess the contributions made by salt bridges to protein stability:[80,85−88] such calculations involve assessing the free energy change due to replacing a pair of oppositely charged residues with identically shaped, uncharged residues, and even in these cases, sensitivity of the results to the value of the protein dielectric has been noted.[80] Finally, one other important PB application, that itself has formed the subject of a number of studies,[66,89,90] is the calculation of pK_a values for ionizable residues in proteins: in effect, what is calculated in such studies is the binding free energy of the *neutral* form of an amino acid to the rest of the protein, relative to the binding free energy of its *charged* form.

In addition to calculating electrostatic free energies, methods have also been developed that allow electrostatic forces on atoms to be computed from a solution to the PB equation.[91] These forces have been successfully combined with an otherwise conventional MM force field to allow dynamics simulations of small molecules (actually using stochastic dynamics methods[47]) in which the effects of solvent are included implicitly.[92] This is feasible for small molecules, for which very small 3-dimensional grids can be used to solve the equation, but for proteins and especially protein−protein systems, which require much larger grids, the current speed of solving the PB equation is insufficient to permit its routine use in any meaningful dynamic simulation: several relatively short simulations have, however, been reported.[92−95]

Because of this, attempts have been made to develop more approximate but extremely rapid techniques for calculating electrostatic interactions. One method that is especially worth noting is the "effective charge" method developed by Gabdoulline and Wade.[96] In this treatment, the electrostatic interaction free energy is described using eq 2, but with the following condition: the electrostatic potential is not resolved for each protein−protein geometry, but instead is solved only once for each protein in isolation. The contribution to the electrostatic free energy from protein 1 is calculated by multiplying its charges by the electrostatic potential due *only* to protein 2. The contributions from protein 2 are obtained by multiplying its charges by the potential due only to protein 1. It is obviously a significant approximation to assume that neither protein affects the electrostatic potential generated by the other, but the method can yield surprisingly accurate energies as long as the charges are scaled or replaced by a set of effective charges that are derived so as to reproduce the protein's own electrostatic potential.[96] The approach works especially well in cases where desolvation effects are relatively unimportant, i.e., for configurations for which the protein surfaces are separated by one or more layers of solvent. As we shall see in a subsequent section, the development of this rapid technique for treating electrostatic interactions has permitted dynamical simulation of the association of proteins in solution.

Nonpolar Interactions

It has been known for many years that the unfavorable hydration free energies of nonpolar solutes such as linear alkanes

are proportional to their accessible surface area.[97,98] The cause of the relationship remains the subject of interesting, ongoing work (e.g., refs 99 and 100). Although it is likely that better approximate methods for treating nonpolar interactions will be developed in the near future, it is safe to say that the surface area approximation is currently the most widely used approach. The approximation amounts to stating that the nonpolar contributions to the hydration free energy of a molecule, ΔG_{np}, can be expressed by the following:

$$\Delta G_{np} = \gamma \text{ SASA} \quad (3)$$

where SASA is the solvent accessible surface area of the macromolecule and γ is a proportionality constant. To estimate the nonpolar hydration contributions to binding of two molecules then, all that is required is that eq 3 be applied first to the complex and then to the isolated partner molecules. The proportionality constant γ has been found by fitting experimental hydration free energies of nonpolar molecules to SASAs to be of the order of 5 (cal/mol)/Å^2.[69] The hydration free energy of a hydrophobic side chain such as valine is therefore around +2 kcal/mol, so complete burial of this side chain in a protein interior is expected to contribute a corresponding amount to the overall folding free energy of the protein.

It is worth noting that drastically different values of γ are often encountered in the literature.[25,101] This is because the energetics of other processes are also often assumed to be linearly related to the SASA; a good example is the free energy of transfer of linear alkanes from their pure liquid state to the aqueous state. The proportionality constant for this process is much larger (~58 (cal/mol)/Å^2) than that for the gas−aqueous phase transfer, because of the additional work that must be done to break the favorable van der Waals interactions in the pure liquid state.[102,103] It has been noted that there are similarities between the transfer of a nonpolar solute from the aqueous state to its pure liquid state, and the transfer of a nonpolar molecule from aqueous solution into a nonpolar binding site in a protein (e.g., ref 103). Although the degree of similarity obviously depends on how closely we consider the protein environment resembles a liquid alkane environment,[104] it has been shown that γ values derived from fitting to this experimental transfer data can be useful for describing nonpolar contributions to binding.[103] It is important to realize, however, that when used in this context, the SASA term is supposed to describe the combined effects of *all* nonpolar contributions to binding—including for example, the van der Waals interactions acting between the two proteins. Since the latter are often described explicitly by the MM force field, care must be taken to avoid double counting of these kinds of interactions.

Some workers have also attempted to use a SASA-based approach for describing the contributions of *polar* groups to hydration.[105,106] Of course, very different γ values are generally required for such groups, since their hydration free energies are favorable (i.e., negative); in fact Eisenberg and MacLachlan's model uses a total of five different γ values, to describe atoms in different chemical environments.[105] Use of SASA terms to describe the energetics of polar and charged group hydration has been criticized (rightly) on theoretical grounds:[107] for example, a halving of the surface area of a charged amino acid need not necessarily result in anything like a halving of its hydration free energy. Although this may not matter too much when considering the hydration of entire proteins due to the fact that the total hydration free energy is a sum of contributions from so many amino acids, a more theoretically sound approach for describing hydration of polar and charged groups is to use electrostatic treatments like the PB equation.[58,107] Nevertheless, SASA models have been used to describe not only the hydration contributions of polar groups, but also *all* their contributions to binding. A model that attempted to correlate the association free energy of a number of protein−protein complexes with SASA terms for example, did produce rather good agreement, when different γ values were used for the nonpolar and polar surface contributions.[25] SASA-based methods can therefore be profitably used if proper consideration is taken of what contributions are included and what are to be explicitly accounted for in other parts of the energetic model.

A very good example of the current state of the art in energetic models for describing protein−protein interactions has been provided in recent work by Kollman's group aiming to conduct the computational equivalent of the alanine scanning experiments referred to earlier.[108,109] Since their method appears to have produced rather good agreement with experimental results, it is worth briefly noting some details of their approach. Starting from the crystal structure of a protein−protein complex, relatively short (~100 ps) conventional MD simulations were performed with explicit water molecules included in order to generate a representative sample of conformations of the complex. The set of structures so obtained were then stripped of their water molecules, and a combined PB/SASA approach was used to calculate the solvation energies of the complex and the separated proteins; the other energetic terms such as internal bonding contributions and van der Waals interactions were provided by the MM force field used in the MD simulations. The entire process was then repeated for the alanine mutants, with their structures being obtained by simply removing the appropriate side-chain atoms, leaving only the methyl groups. This ensures that the only difference between the structure of the wild-type (i.e., original) and mutant proteins is the absence of the rest of the side chain, thus minimizing noise in the calculations. Obviously, however, this method will not work for other mutations that one might wish to examine in future, such as the inclusion of a side chain that contains more atoms.

One point that is particularly striking is their successful reproduction of the energetic effects of mutations of charged amino acids. This is especially notable given the fact that in their PB calculations a protein dielectric constant of 1 is used. This is the correct approach to adopt when multiple conformations are being examined, since it assumes that the dielectric response of the protein can be accurately captured by the distribution of structures and therefore need not be implicitly modeled in the electrostatics calculations. Nevertheless, it also means that the desolvation energies that accompany association are very large. This can be seen from the fact that in their calculations, the mean change in the PB solvation free energy that results from the mutation of one of the charged amino acids is almost 70 kcal/mol (see Table 2 of ref 108). This change is compensated by similarly large changes in the mean Coulombic interaction energy. That near perfect cancellation of the electrostatic solvation and interaction terms is obtained is therefore notable, especially in view of the fact that the variations of these terms within the set of structures are often very large (standard deviations of ~25 kcal/mol). As noted above, it is partly the tremendous sensitivity of results to the value of the protein dielectric constant that has up to now limited applications of the method; the results obtained by Kollman's group are therefore either the result of an excellent parametrization of the method, a fortuitous cancellation of opposing terms, or (most interestingly) a consequence of averaging results over a large number of structures. Obviously, the only way to

establish which interpretation is correct is to apply the method to other systems.

Protein Docking Problem

In discussing computational methods for describing protein−protein interactions, it is impossible to ignore the so-called "protein docking" problem, since for structural biologists, the solution of this problem represents something of a search for the Holy Grail.[110−112] Put simply, the aim is to correctly predict the structure of a protein−protein complex, given only the structures of the separated partners. The latter clause is important, since predicting the structures of proteins (given only their amino acid sequences) represents an even more ambitious goal of computational biology (for a review see ref 113).

The importance of the docking problem is easy to demonstrate. The tremendous advances in genome sequencing have provided the research community with the amino acid sequences of thousands of gene products (i.e., proteins) whose three-dimensional structures are as yet unknown. The realization that continuing to solve all of these crystal structures through traditional means would take years, has resulted in the advent of the field of *structural genomics*: large-scale, highly automated crystallization initiatives aimed at solving as many structures as possible in as short a time as possible.[114] In passing, it is worth noting that computational methods aimed at predicting protein structure have already proven useful in guiding such studies by providing an indication of which proteins are more likely to yield structures that are truly novel (i.e., which are not closely similar to structures of already solved proteins).[115] But solving the structures of the proteins themselves is not always the final aim: after all, many proteins perform their biological function not in isolation but in interacting with other macromolecules.[116] It is the understanding of these regulatory interactions that is often the more important goal, and to achieve this fully requires knowing the structure of the complex. Unfortunately, solving the structures of protein−protein complexes is an even more difficult undertaking than solving the structures of the component proteins themselves: many complexes are simply not amenable to crystallization, even if their constituent proteins are.

It is for this reason that interest has been spurred into trying to solve the problem by computational means. Given that the problem is also of intrinsic academic interest, it is no surprise that a number of algorithms have been developed, but although substantial progress has been made since the first study more than 20 years ago,[117] it is fair to say that it remains essentially unsolved (see, for example, refs 110 and 111). At first sight this might seem a little puzzling; after all, one might ask how many different ways there are of fitting two proteins together? Certainly, we can expect that the problem should be many orders of magnitude easier to solve than the related protein folding problem, simply because the orientational possibilities are so much fewer. At its most basic level, i.e., if we treat both proteins as rigid bodies, there are only six degrees of freedom (three translations and three rotations of the second protein relative to the first). The protein folding problem on the other hand has at least 3^{2N} degrees of freedom, where N is the number of amino acids in the protein (assuming conservatively that each amino acid in the chain can adopt only 3^2 possible conformations). In fact, if the problem was as simple as a 6-dimensional search, it might already be solved: computational advances in the past few years have meant that conducting a search through the billions of possible relative orientations of two proteins is in some sense no longer a problem. So why is the docking problem itself not solved?

We can point to two major factors. First, an energetic function is needed in order to select the most favorable orientations. Obviously, there will be a tradeoff between the complexity of the energy function used, and the speed of conducting an orientational search: energy functions that are rapid to compute will allow many more potential protein−protein structures to be tested but may be less adept at discriminating between good and bad structures. Although we have noted in the previous section that relatively elaborate energetic models are available for treating protein−protein interactions, most docking algorithms use much simpler scoring functions that aim to select structures that maximize the complementarity of the two protein surfaces (e.g., ref 118). This approach simply reflects the prevalence of the point raised earlier, that protein−protein complexes are often characterized by a very high degree of surface complementarity. Given the previous discussion, it should be clear that equating the strength of interaction with the amount of buried surface area makes sense for interactions between hydrophobic side chains. For describing the interaction energies of polar or charged surfaces, the relationship is less easy to justify, although not completely unreasonable, as long as different proportionality constants are used.[25] However, most existing algorithms do not make any distinction between polar and nonpolar regions, so buried surface area is treated as a uniformly favorable contributor to binding, independent of the nature of the amino acids involved.

In addition to its conceptual simplicity, the other advantage of adopting a surface complementarity function is that it can be formulated as a convolution and so can be readily calculated by fast Fourier transform methods (FFT; ref 119). As a result, the generation of potential protein−protein geometries can be conducted using a grid-based approach, in which a 3-dimensional grid is placed over one of the proteins, and the center of the other protein is then placed in turn at each of the grid points. For each grid point, hundreds or thousands of different orientations of the second protein are obtained by rotation around the three Cartesian axes. The current state of the art for such calculations is such that during a typical run of the FFT program DOT,[120] a total of $128 \times 128 \times 128 \times 51000 = 1.07 \times 10^{11}$ structures can be generated and their interaction energies calculated. It has also been shown that electrostatic interactions can be formulated as convolutions,[121] as long as the interaction energy is approximated (as in the effective charge method) by multiplying the charges on each protein by the electrostatic potential generated at each charge by the other protein. Again, from the previous discussion, we know that such an approximation can work very well when charges are scaled appropriately,[96] but it does not take into account the unfavorable consequences of "electrostatic desolvation". Since the latter effects can have major energetic and even structural consequences for macromolecular interactions (e.g., ref 94), particularly in highly charged systems, their inclusion is likely to be essential to a complete energetic formulation.

Although it should be obvious that fuller energetic functions should be capable of better discrimination of potential structures, it is not at all easy to see how they can be made sufficiently rapid to include in an exhaustive search. Obviously, one way around this problem is simply to use the more elaborate energetic model only after an initial grid search procedure has been used to select structures with high degrees of surface complementarity.[122−126] This approach has been tried with a number of rather complete energetic models that take account of factors such as electrostatic desolvation, even in one case including the entropic penalty of restricting side-chain freedom.[122−126]

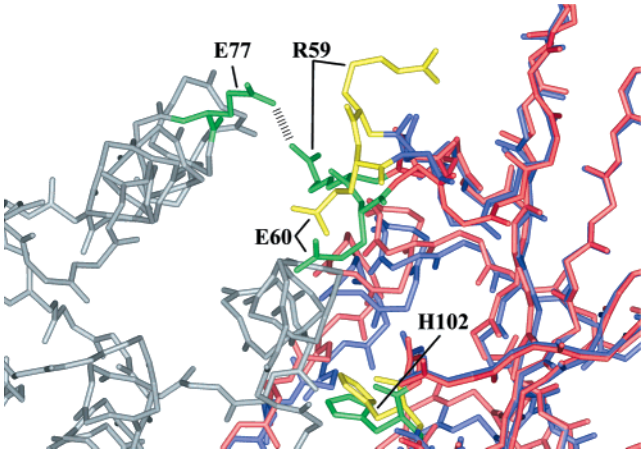


Figure 4. Example of a minor conformational change accompanying association of barnase with barstar. The gray trace represents the backbone of barstar. The blue trace represents the backbone of barnase when crystallized in isolation; the red trace is the backbone when crystallized with barstar present. Several amino side chains undergo large positional shifts when changing from the unbound (yellow) to the bound (green) state.

There is, however, a clear danger with such an approach, which is that if the most favorable binding orientation involves only a comparatively small interface, it is likely to be overlooked, and so will not be included in the list of structures subjected to the full energetic calculations. This is not as absurd a claim as it might at first sight appear: there are a number of dimeric proteins for example whose monomer−monomer interfaces are so small that they would almost certainly be overlooked in a search based solely on buried surface area considerations.[127]

A second reason the protein docking problem remains currently unsolved is that even if the energetic function used in the search were adequate, properly specifying a protein−protein complex involves significantly more than 6 degrees of freedom. Structural changes accompany all protein−protein association events, even if they are only minor side-chain reorientations (Figure 4; refs 128 and 129) and models that treat proteins as rigid bodies obviously neglect this aspect. Despite this, the rigid body approximation can work surprisingly well, even when the structures are treated at very low resolution,[118] and it is often straightforward to reassemble the crystal structure of a complex from its separated components. But this is not the same as solving the complete problem: reassembling a complex is a much simpler undertaking because the amino acid side chains are already positioned in their bound state and so fit together very snugly. Solving the general problem on the other hand, requires the use of free, independently solved structures for the partners, whose side chains will certainly not be optimally arranged. In general, it will therefore be far more difficult to correctly predict the structure of the complex.[130] A crude way of dealing with the flexibility problem is to allow some softness to the protein's surface, so that some degree of steric overlap of side chains is allowed without resulting in tremendously unfavorable energies.[111] A general solution will, however, certainly require flexibility to be properly introduced. Unfortunately, this raises very severe computational difficulties: in particular, it is not easily reconciled with FFT-based grid searches.[111,129] Nevertheless, Vajda and co-workers have very recently shown that very good discrimination between docked structures can be achieved if energy minimization of side chains is allowed.[124]

Although the above discussion covers the most fundamental difficulties, there is an additional problem that has so far received comparatively little attention, namely the potential role played by explicit solvent molecules. The use of surface area descriptions and continuum solvation methods implicitly suggests that in most researchers' opinions, a correct answer to the docking problem can be arrived at without consideration of the molecular nature of the solvent. But, as we have noted above, water molecules are identifiable in almost all protein−protein interfaces, and a theoretical analysis has suggested that these interfacial waters contribute as much as 25% to the overall strength of a typical protein−protein interaction.[131] The lack of an explicit accounting for water molecules in docking calculations may therefore conceivably hamper identification of the correct structure. This is not to say, however, that every complex with a large number of water molecules at the protein−protein interface will evade successful prediction. In fact, in a challenge set by a group of crystallographers,[132] the crystal structure of a complex between two proteins BLIP and TEM1 was successfully predicted by all 6 groups participating, even though the complex contains a total of ∼20 water molecules in the interface. Although this result indicates that in certain cases successful docking can be accomplished with extant models, it was rightly pointed out by Shoichet and Kuntz[133] that the unusually large buried SASA of the particular complex examined was almost certainly a major factor contributing to the success of the docking algorithms. As these authors noted, even when the correct answer was obtained, specific interactions such as hydrogen bonds and salt bridges observed in the crystal were not reproduced. An explicit treatment of solvent will clearly not be required in all cases then, but there is no guarantee that this will carry over to *all* cases.

Finally, it is worth noting that the search problem can in many cases be made somewhat simpler by incorporating additional information. In the docking challenge discussed above, for example, several groups used the fact that the inhibitor is likely to occlude the active site of the enzyme as a screen for eliminating unlikely structures. In other cases, it may already be known from biochemical studies such as alanine scanning experiments which amino acids are involved in the interface. Sometimes potential binding sites on the surface of a protein can be identified by an analysis of the surface's physicochemical properties.[134] Finally, sequence analysis methods may be of use for providing at least first guesses for the localization of protein−protein interaction surfaces, since the residues involved in binding are likely to be subject to evolutionary conservation.[40] Obviously, there will be limits to the accuracy of such predictions, since an interface is after all, by definition, a function of two protein surfaces not one.

Weak Interactions

The protein docking problem forms the most explicit and rigorous test of the accuracy of energetic models of protein−protein interactions, since the test is more or less unambiguous: the docking method (and its associated energetic model) either predicts the correct structure or it does not. However, not all protein−protein interactions can be characterized structurally since many complexes prove next to impossible to crystallize, often because the interactions are inherently weak. There is no need to suppose, however, that there should be a direct correlation between the strength of a protein−protein interaction and its physiological importance. In other words, while we may find it convenient to ignore weak interactions because of the difficulties involved in studying them, there is no real justification, other than expediency, to do this. In fact, many important protein−protein interactions are known to be

surprisingly weak, with good biological examples being provided by the transient interactions formed between electron-transfer proteins,[135] and the weak interactions involved in holding multienzyme complexes together in vivo (e.g., ref 136). Interesting biophysical manifestations of weak or nonspecific interactions are lowered apparent diffusion coefficients of macromolecules in the cytoplasm or other intracellular compartments (e.g., refs 7−9).

The description of weak interactions represents a considerable challenge for theoretical methods precisely because there is so little useable experimental data with which to test the models. Structural information from crystallography is as we have noted, more or less nonexistent, and the other main tool of structural biology, NMR, is equally ineffective for cases where association is comparatively weak. In addition, for some macromolecular systems even more basic information is often missing: in the case of most multienzyme complexes for example (with one or two notable exceptions) even the stoichiometries of the various component enzymes are unknown. Nevertheless, there is at least one experimentally accessible quantity that can be readily measured for weak *two-body* interactions between macromolecules. This is the second virial coefficient B_{22},[137] formally defined as follows:

$$B_{22} = \frac{-1}{16\pi^2 \mathrm{MW}^2} \int (\mathrm{e}^{-\Delta G_{\mathrm{inter}}/kT} - 1)\, \mathrm{d}\Omega \qquad (4)$$

where $\Delta G_{\mathrm{inter}}$ represents the interaction free energy (potential of mean force) acting between two protein molecules in solution, MW is the molecular weight of the protein(s), and the $\mathrm{d}\Omega$ indicates that the integral is over all possible relative positions and orientations of the two molecules. Although B_{22} can in principle be measured for mixed protein solutions, we will consider here only single component protein solutions, for which $\Delta G_{\mathrm{inter}}$ describes the interaction between two *identical* macromolecules. In words, B_{22} quantifies the effects of two-body interactions in causing deviations from a solution's ideal behavior. Obviously, at the simplest level, two proteins can be considered to interact solely through steric means: two molecules cannot occupy the same position in space at the same time. These (unfavorable) steric interactions result in what is generally called an "excluded volume" term and in the absence of other interactions cause B_{22} to take on positive values. Other interactions, can of course be either favorable or unfavorable, with the former shifting B_{22} toward negative values. The main advantage of B_{22} as a gauge of protein−protein interactions is its ease of measurement: a variety of experiments, such as osmotic pressure, static and dynamic light scattering, and sedimentation equilibrium measurements have all been used (for a discussion, see ref 5).

The second virial coefficient is sensitive to both the ionic strength and the pH of the solution: both effects are usually rationalized simply in terms of monopole−monopole electrostatic interactions. A common observation for example is that B_{22} becomes more positive in value as the pH moves further away from the protein's pI (i.e., the pH at which it has a zero net charge). This can be interpreted simply in terms of the unfavorable monopole-monopole interactions becoming stronger as the net charge on the protein increases in magnitude. Similarly, another common result is to find that increasing salt concentration is accompanied by decreases in B_{22}; this is interpreted as being due to increased screening of the repulsive electrostatic interactions. Although most common, these are not the only trends observed. For example, in the case of chymotrypsinogen, it has been found that B_{22} actually increases with increasing salt concentration at pH 7.[5] This result can only be explained by the presence of favorable dipolar (or possibly higher order) electrostatic interactions that become progressively screened at higher salt concentrations.

Many theoretical approaches to the calculation of B_{22} have been reported, but until recently, such calculations utilized highly simplified (spherical) representations of protein structures. Hard-sphere-only models are sufficiently simple to deal with that analytical expressions for all virial coefficients up to the seventh order can be derived,[138,139] but they are limited because the protein's radius is the only parameter that can be adjusted to fit experimental results. In cases involving strong attractive interactions for example, fitting to experimental B_{22} results with hard-sphere-only models has even required *negative* protein radii;[140] this is clearly a physically absurd result that must place severe limits on the utility of such a model. More elaborate treatments have modeled the interactions using DLVO type potentials, which retain hard-sphere terms, but augmented with electrostatic interactions treated as a sum of monopole-monopole, monopole−dipole and dipole−dipole contributions and van der Waals interactions modeled using Hamaker type formulations. Even with these kinds of models, there is a clear disadvantage: even though they can provide a phenomenological description of experimental results,[5] they are clearly of no use in providing any kind of structural interpretation of B_{22} values.

This of course would not be such a problem if B_{22} were really only a nonspecific measure of protein−protein interactions, which did not depend in any meaningful way on the atomic nature of the interactions. This, however, is not the case, as has been demonstrated in a series of papers from Lenhoff's group pioneering the use of atomically detailed models in calculations of B_{22}. In their first use of such a model, aimed at simply calculating the excluded volume contribution to B_{22}, the result obtained using an atomic model was found to be some 40% larger than would have been estimated using a spherical model of the same volume.[141] This increase was attributed to the roughness of the protein surface. In a more recent work, which also provides an excellent review of the issues encountered in conducting B_{22} calculations, an energetic model was developed to include both electrostatic and van der Waals interactions in the calculations.[142] Electrostatic interactions were described using a PB continuum model, but with the calculations accelerated by the nonnegligible (and far from ideal) approximation that the low dielectric regions occupied by the proteins are spherical in shape. Short-range interactions, encompassing hydrogen bonding, van der Waals interactions and hydrophobic effects were dealt with using a hybrid Hamaker/Lennard-Jones treatment with the former model being used to describe interactions between atoms separated by more than 6 Å, and the latter being used for all closer interactions. This same model has recently been further parametrized[143] in an attempt to reproduce the protein−protein binding free energies tabulated by Horton and Lewis.[25]

The key finding made by Neal et al.[142] was that the largest contributions to the calculated B_{22} values come from configurations in which the two proteins are in highly complementary orientations. A glance at eq 4 shows that this results from the fact that disproportionately large contributions to the integral are made by energetically favorable structures (i.e., structures for which $\Delta G_{\mathrm{inter}}$ is strongly negative). This result suggests that even though B_{22} values can often be explained in gross terms by treating proteins as interacting monopoles and dipoles, they may be sensitive to specific atom−atom interactions. Two recent results support this interpretation. The first is again from

Lenhoff's group, who have shown that substantial changes in B_{22} values can result from mutations of a single amino acid side chain, even when this mutation does not result in a change in the net charge of the protein.[144] The second comes from our own work aimed at extending atomically detailed B_{22} calculations in order to develop a transferable energetic model capable of providing a general description of weak protein−protein interactions.[145] In our case, evaluation of the integral of eq 4 was performed using a grid-based approach more or less identical to that discussed above in the context of protein docking: the first protein was placed at the center of the grid and the second molecule was placed in a number of different orientations at all points on the grid. Although the energetic model we use is quite detailed (incorporating explicit terms for electrostatic desolvation) it is not really any more successful at reproducing B_{22} values at the extremes of pH than is a much simpler DLVO type model.[5] By performing more complete energetic calculations on the 200 most favorable structures identified in the grid search, we have obtained strong evidence that the difficulties in fitting experimental results at extreme pHs are due to a neglect of the fact that ionizable amino acids can change protonation state on complexation. Since these protonation state changes result from the intimate interaction of specific amino acid side chains, it is clear that these effects will be impossible to capture in more structurally simplified models.

Protein−Protein Association Kinetics

The applications discussed so far have emphasized the thermodynamic aspects of protein−protein interactions: both virial coefficient calculations and docking methods combine energetic models similar to those discussed above with grid-based searches of potential protein−protein geometries. In contrast, in this section we discuss the use of the same energetic models to simulate the *dynamics* of protein−protein interactions, since such applications represent one of the most successful applications of computer simulation techniques of atomic detail.

The kinetics of macromolecular association events are of interest for a number of reasons, but perhaps the most vivid illustration of their potential biological importance comes from noting that a number of enzymes have evolved to such an extent that the rate-limiting step of their catalytic cycle is their association with their substrates. Examples of such enzymes, which are termed diffusion-limited (or "perfect") enzymes, include superoxide dismutase (SOD), triose phosphate isomerase (TIM), and acetylcholinesterase (AChE).[146] Their extreme efficiency is no accident: SOD catalyzes the destruction of the potentially harmful superoxide radical, while AChE is responsible for destroying the neurotransmitter acetylcholine so that a neuron or muscle fiber, once stimulated, does not continue to fire.[147] Both of these enzymes are therefore clearly under evolutionary pressures to act rapidly. Because association is the process that places an upper limit on the overall efficiency of these enzymes, simulations that can address this process may, in addition to providing insight into the factors controlling the process, also suggest ways to accelerate (or otherwise modulate) the process so as to produce more efficient enzymes.

We have already noted that molecular dynamics simulations of proteins in which solvent molecules are included are currently limited to a time scale of at most hundreds of nanoseconds. This is too short for routine observation of macromolecular association events, so again the continuum solvent approximation is adopted allowing explicit accounting of water molecules to be avoided. Of course, in addition to their effects on protein−protein interaction thermodynamics, water molecules exert other, purely dynamic effects. Stochastic collisions with water molecules, and the viscous drag effects that they exert both affect the motion of molecules in solution, with the result being that solutes exhibit Brownian motion. These dynamic effects can be modeled in an implicit fashion with the technique of Brownian dynamics (BD) using a combination of a suitably scaled diffusion coefficient for the molecule of interest (to account for solvent viscosity), and the inclusion of a random force component (to model the effects of collisions with solvent molecules). The basic algorithm of BD discussed here is that due to Ermak and McCammon:[148]

$$\mathbf{R}(t+\Delta t) = \mathbf{R}(t) + \frac{D\Delta t\mathbf{F}}{kT} + \mathbf{S} \tag{5}$$

where $\mathbf{R}$ is the position of the particle, D is its diffusion coefficient, $\mathbf{F}$ is the force acting on it due to interactions with other particles, and $\mathbf{S}$ is a random displacement vector of zero mean. Δt is the time step, k, Boltzmann's constant, and T, the temperature. The algorithm provides a straightforward recipe for continually updating the positions of Brownian particles so as to simulate their dynamic behavior. Computational speed is achieved partly through the fact that the positions of water molecules are not explicitly modeled, and partly because much larger time steps (Δt) can be used in BD than in MD. In the former, time steps of 1 ps are common, representing a 1000-fold increase over the 1 fs time steps generally used in MD simulations of condensed phase systems. The Ermak−McCammon algorithm, or other close variants, has been used extensively in simulations of colloidal solutions, in which interparticle interactions are treated using DLVO or Yukawa-type potentials (e.g., refs 149 and 150). Such simulations have given insight into changes in diffusion coefficients that result from increasing protein concentrations as well as providing views of the onset of glass transitions (see ref 151). Of more relevance here however is the fact that BD simulations have also been conducted using atomically detailed protein models, initially for simulating the association of small molecule substrates with enzymes, but also more recently for simulating actual protein−protein association events.

The method by which association rate constants are calculated from BD simulations is straightforward, and since it has been reviewed several times,[75,152] we limit our discussion here to the essential aspects. A series of simulations are performed, in each case with the two associating molecules initially separated by some distance b. The positions of the particles are updated using eq 5 until either the two molecules associate, or they reach some much further degree of separation (q), at which point they are assumed to have escaped from each other (Figure 5). How one defines association is an interesting (and still not completely resolved) question (see, e.g., ref 153): generally, this is done by monitoring the distances between certain pairs of atoms known to be adjacent to each other in the final complex. Association is assumed when one or more of these atom pairs come within a specified distance (usually ∼6−7 Å). This is equivalent to assuming that once the atoms reach this separation that subsequent association is completely assured. By performing many (thousands of) simulations, statistically robust estimates of the probability β, that association will occur can be calculated. Once this has been obtained, it can be combined with Smoluchowski's analytical result[154] for the rate constant for two

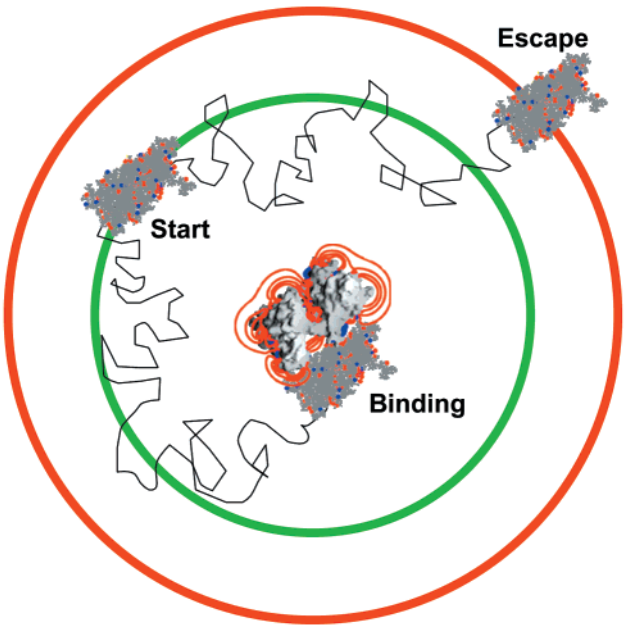


Figure 5. Schematic illustration of the use of Brownian dynamics simulations for calculating association rate constants. At the start of each simulation, the two proteins are separated by a distance b (green surface). Diffusion of one of the proteins is then simulated until it either binds to the other protein or escapes to some larger distance q (red surface). See text for details.

proteins achieving the initial separation (k_b), to give an overall rate constant for the association of the two molecules, k_{assoc}:

$$k_{assoc} = k_b\beta \tag{6}$$

In many discussions, an extra term appears in the denominator of eq 6 because simulations should technically be terminated only when the proteins achieve infinite separation, not when they become separated by the distance q. For simplicity we have omitted this term. Smoluchowski's analytical expression for the rate constant for two spherical particles achieving a separation b is straightforward: $k_b = 4\pi Db$, where D is the relative translational diffusion constant of the two molecules. In the case of small molecule−enzyme association events, the relative diffusion coefficient is dominated by the diffusion coefficient of the small molecule since this is generally one or 2 orders of magnitude larger than that of the much larger enzyme. If diffusion coefficients are not available from experimental measurements (see Creighton[155] for a representative list for proteins), they can be estimated either using the Stokes−Einstein relations:

$$D_{translation} = \frac{kT}{6\pi\eta R} \qquad D_{rotation} = \frac{kT}{8\pi\eta R^3} \tag{7}$$

where k is Boltzmann's constant, T is the temperature, η is the solvent viscosity, and R is the radius of the macromolecule, or they can be estimated using more rigorous hydrodynamic theories that take explicit account of the 3-dimensional geometry of the molecule (see, e.g., ref 156).

Simulations of enzyme−substrate association events in particular have been extremely successful, and have allowed the dependence of reaction rate constants on ionic strength and amino acid mutations to be rationalized.[157−162] Common to these studies has been the finding that association is dramatically accelerated by favorable electrostatic interactions between the two associating molecules, even when they carry net charges

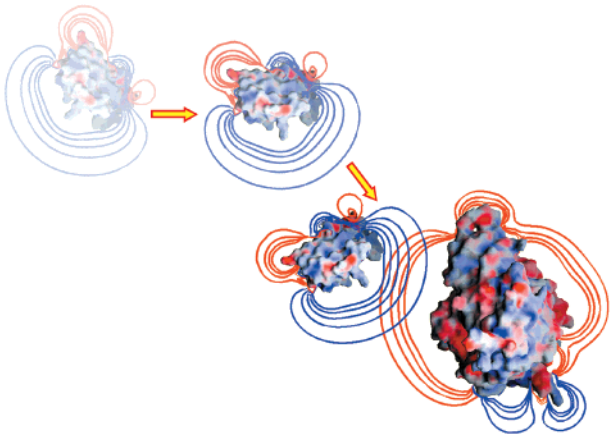


Figure 6. Schematic illustration of the association of fasciculin (at left) with acetylcholinesterase. Blue and red contour lines indicate regions of positive and negative electrostatic potential, respectively.

of the same sign. One especially good demonstration of the potential power of these studies was provided by work in which simulations were employed to rationally design a faster version of the enzyme SOD which was then synthesized and tested in the laboratory.[158] The uses of atomically detailed BD models in simulating protein−protein association events are fewer, mainly because only recently has sufficient computational power become available (see ref 163 for a review and refs 164−170). The first such studies were conducted by Northrup and co-workers using their MacroDox program, to simulate electron-transfer rates in cytochrome systems.[164] More recently, other systems that have been examined are the association of barnase with its inhibitor barstar[168,171] and acetylcholinesterase with the snake venom toxin fasciculin (Figure 6 and refs 161 and 169). Experimentally it had been shown that in these cases, association is a diffusion-limited process (i.e., conformational rearrangements accompanying binding are *not* rate-limiting) and, as in the case of the enzyme−substrate systems, is dramatically accelerated by electrostatic interactions. The former fact can be ascertained by performing experiments in solutions containing different concentrations of otherwise inert solute molecules that increase the solution viscosity; the latter fact is generally demonstrated by simply altering the ionic strength of the solution by changing the concentrations of added salts.

BD simulations using the program SDA[168] successfully reproduced these experimental results. In the simulations, the systematic forces acting on the proteins were assumed to be entirely electrostatic, and were described using the effective charge model discussed earlier. In the case of barnase−barstar, the simulations reproduced both the ionic strength of the association rate constants and their dependence on single (and one double) amino acid mutations in the associating proteins. The calculated rate constants spanned 2 orders of magnitude in 50 mM NaCl. In the case of acetylcholinesterase−fasciculin it was shown that this accelerated association occurs even when the electrostatic interactions do not strongly stabilize the complex; in other words, kinetics can be accelerated even if the thermodynamics of binding is not dramatically improved. More recently, the same methodology has been used to demonstrate why actin polymerization occurs more rapidly at the so-called "barbed" end than at the "pointed" end of actin filaments.[170]

It is worth noting that a much simpler but highly effective route to calculating association kinetics has also been developed. Selzer and Schreiber[172,173] have built upon the work of Zhou and co-workers[174] to show that relative association rate constants

of mutant proteins are very highly correlated with their electrostatic interaction energy in the final complex. The power of this relationship has been demonstrated in the most explicit and elegant way, by the rational design and testing of a number of faster associating proteins.[173] There are, not surprisingly, likely to be limits to such an approach, the clearest being in the case where there is a substantial degree of nonspecific binding to other parts of one of the partner molecules. A very good example of such a phenomenon has been provided by studies of restriction enzymes such as EcoRV. A series of clever experiments have shown that association of these enzymes with their recognition sites on DNA is faster in long stretches of DNA than in short stretches.[175,176] The reason for this effect is that the enzyme first binds nonspecifically to (any point on) the DNA and then diffuses along the double helix to its specific binding site. In doing so, the enzyme in effect converts a 3-dimensional search problem (how to find a small binding site) into a 1-dimensional search, which is much more rapid to perform.[177,178] Since accelerated association in this way is dependent on binding to sites different from the eventual target, it is difficult to see how this could be captured in approximate models that consider only the final binding site.

Finally, interesting simulation work has also been reported aimed at establishing factors responsible for fast association of proteins in cases where electrostatic interactions are not important.[179,180] Camacho et al.[179] have recently reported an interesting updated study of the problem originally considered by Northrup and Erickson,[180] but using a much more detailed and realistic energetic model, consisting of a combination of atomic desolvation parameters derived by Zhang et al.[181] together with a simple approximation to the electrostatic interaction. A particularly interesting feature of their BD simulations was the demonstration that a uniformly distributed (i.e., centrosymmetric) interaction potential could not fit the experimental rate association rate constants without simultaneously predicting unrealistically long association times ($\sim$1 μs). Since centrosymmetric potentials are almost always used in DLVO-type modeling, this may turn out to be a rather important result. When a (presumably) more realistic interaction potential was used, rapid association was observed even with short average lifetimes of nonspecific binding of $\sim$4 ns. Accelerated association was obtained even in cases where the interaction energy averaged over all possible structures of the encounter complex was actually repulsive: the existence of only a small number of attractive protein−protein geometries was shown to be sufficient to cause dramatic increases in the association rate.

Discussion

Our main goal in writing this work has been to provide a broad overview of atomically detailed simulation methods in a way that is accessible to people from other fields. The reason for emphasizing this point is that there are significant differences between atomically detailed models and the DLVO and other related models that are probably more familiar to the readership of this journal (for a review see ref 182). Of course, the different methods used reflect the specific interests of the researchers, and this difference in perspective is reflected even in the meaning of the phrase "protein−protein interaction". To workers in the colloidal field, interaction generally describes the relatively long range influence of electrostatic interactions on structural and diffusional properties of macromolecular solutions. What goes on at contact is of almost no interest, or at least it is considered sufficient to model an interacting protein or colloidal particle as a sphere of appropriate radius and charge with a suitably parametrized term to describe van der Waals interactions. Certainly the physicochemical nature of the protein−protein interface is never discussed. To crystallographers of protein−protein complexes, and biologically minded researchers on the other hand, interaction implies a much more intimate association, generally of a quite specific nature. Two proteins are only considered to interact if they form stable, isolatable complexes and even nonspecific interactions are still assumed to involve a substantial degree of contact between the partner proteins. Undoubtedly, workers in both fields can benefit from considering other viewpoints: biologically minded researchers for example would do well to note that meaningful protein−protein interactions can operate over large distances (beyond contact) and still have both thermodynamic and kinetic consequences. The question of how much these effects operate in a biological setting should form one of the more interesting avenues for research in the coming years.

One aspect that will have been clear from this work is that within the area of atomically detailed models there are likely to be different levels of approximation that are acceptable depending on the specific application of interest. For example, it is a safe bet that a general solution to the protein docking problem, assuming that it does not require any new physics to be discovered, will certainly require the inclusion of conformational flexibility in the proteins. At the very least, this will involve allowing side chains some freedom of movement, but more generally it will also require backbone flexibility. The question of how to efficiently incorporate these features into existing protein docking algorithms remains unresolved at the present. On the other hand, conformational flexibility may be a less pressing requirement for a general description of weak protein−protein interactions. Crystallographic analyses of nonspecific interactions in protein−DNA complexes indicate that complexation is driven more by electrostatic interactions between charged groups than by surface area burial, although the latter is not insignificant.[183] Similarly, it is usually (though not always) straightforward to separate biologically important oligomerization contacts in crystal structures from nonspecific crystal contacts, simply on the basis of the amount of buried surface area.[184,185] We can therefore be relatively confident that the utility of rigid body models to the simulation of weak protein−protein interactions should be considerably greater than to strong protein−protein interactions. This of course leaves open the question of how strong an interaction has to be in order to be considered a strong interaction!

A further example of the way the requirements of an energetic model may differ depending on the application of interest comes from considering the effects of pH on protein−protein interactions. We have indicated in the section entitled "Weak Interactions", that the calculation of second virial coefficients at the extremes of pH is likely to be hindered by problems due to changes in protonation states accompanying association. It seems likely that a similar effect may contribute in part to the need for charge renormalization in applications of the DLVO model to colloidal interactions (see, e.g., ref 151). Ionization state changes will be especially likely at pH ranges that straddle the pK_a values of the ionizable groups in proteins. Again, although this will be of interest to biophysicists, it is much less likely to appear relevant to biologists whose interest is primarily focused on the physiological pH range around 7. At this pH, changes in protonation states will probably be less frequent because of the paucity of amino acids that titrate: only histidine and N-terminal amino acids. This is not to say, however, that the role of

ionization state changes in determining association equilibria is unimportant biologically. It should be remembered that several biological interactions have been shown to be dependent on the ionization state of certain amino acid side chains,[186] and cases are known where a change in pH can trigger important physiological events such as the fusion of viral and cell membranes.[187]

Although we have emphasized the use of atomically detailed models, it is of course worth noting that in many settings, the use of DLVO type models provides a host of benefits. In particular, the speed of evaluating the interaction energetics and forces means that in comparison to atomically detailed models, many more particles can be simulated for much longer periods of time. Although many physical observables may not be strongly dependent on structural details of the interacting molecules, any desire to establish particular amino acids as being responsible for modulating interactions will obviously remain unfulfilled using such structurally simplified models, except in the relatively trivial cases where the role of the amino acid in question can be described in terms of its effects on the net charge of the particles. It is to be hoped therefore that the more detailed models discussed here will find use in simulating many-particle systems before long, especially as accessible computational power continues to increase at its current rate. For example, given the success of effective charge methods in simulating association kinetics of binary systems, it is not unreasonable to hope that the same techniques might find similar utility in studying the kinetics of ternary (or larger) systems. A useful first step in this direction would be to demonstrate that pairwise interactions described under the effective charge approximation are also capable of accurately reproducing the electrostatic interaction energies of many-body systems. Beyond this, the next major potential problem to be faced may simply be statistical in nature since obtaining accurate estimates of even bimolecular association rate constants can require many thousands of simulations.[168]

Having said that, it is of course worth questioning just how interesting association kinetics are. The fact that association of proteins can be accelerated by favorable electrostatic interactions, for example, ceases to be particularly interesting after it has been demonstrated several times. Of course, there are as-yet unidentified phenomena that may emerge from the complex behavior of many-body systems,[188] and there is no shortage of biological systems and questions that are worth studying. The major problem is how to reduce them into terms that can be readily simulated: the increase in complexity in moving from studying bimolecular association events to protein diffusion in crowded biological media for example is tremendous. Despite this, there are grounds for optimism, in large part because of the relatively recent development of experimental techniques (usually based around the use of fluorescent proteins) for investigating macromolecular diffusion in the cell. In many senses it is fair to say that the questions raised by such studies are often more abundant than the answers they provide.[7−9] This area therefore appears particularly ripe for a useful combination of theoretical and experimental approaches. To our knowledge, the only truly meaningful attempt to simulate a model of macromolecular diffusion in the cytoplasm has been the extremely interesting work of Bicout and Field.[189] This paper seems unfortunately to have been largely ignored by the research community. Nevertheless, in an indication that at least some groups see the potential utility of combined approaches, one experimental group has recently reported simulations in which the effects of the 3-dimensional geometry of subcellular compartments on macromolecular diffusion were modeled.[190] In the future, these kinds of studies are likely to become more frequent.

Acknowledgment. This work was supported by grants to J.A.M. from the National Science Foundation, the National Institutes of Health, and the San Diego Supercomputer Center.

References and Notes

(1) Boriack-Sjodin, P. A.; Margarit, S. M.; Bar-Sagi, D.; Kuriyan, J. *Nature* **1998**, *394*, 337.
(2) Andel, F.; Ladurner, A. G.; Inouye, C.; Tjian, R.; Nogales, E. *Science* **1999**, *286*, 2153.
(3) Guillet, V.; Lapthorn, A.; Hartley, R. W.; Mauguen, Y. *Structure* **1993**, *1*, 165.
(4) Welch, M. D. *Trends Cell Biol.* **1999**, *9*, 423.
(5) Velev, O. D.; Kaler, E. W.; Lenhoff, A. M. *Biophys. J.* **1998**, *75*, 2682.
(6) Price, W. S.; Tsuchiya, F.; Arata, Y. *J. Am. Chem. Soc.* **1999**, *121*, 11503.
(7) Seksek, O.; Biwersi, J.; Verkman, A. S. *J. Cell Biol.* **1997**, *138*, 131.
(8) Elowitz, M. B.; Surette, M. G.; Wolf, P.-E.; Stock, J. B.; Leibler, S. *J. Bacteriol.* **1999**, *181*, 197.
(9) Dayel, M. J.; Hom, E. F. Y.; Verkman, A. S. *Biophys. J.* **1999**, *76*, 2843.
(10) Verweey, E. J. W.; Overbeek, J. Th. G. *Theory of the stability of lyophobic colloids*; Elsevier: New York, 1948.
(11) Bowen, W. R.; Mongruel, A. *Colloids Surf. A* **1998**, *138*, 161.
(12) Janin, J.; Chothia, C. *J. Biol. Chem.* **1990**, *265*, 16027.
(13) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13.
(14) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 121.
(15) Tsai, J.; Lin, S. L.; Wolfson, H.; Nussinov, R. *J. Mol. Biol.* **1996**, *260*, 604.
(16) Stites, W. E. *Chem. Rev.* **1997**, *97*, 1233.
(17) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177.
(18) Xu, D.; Lin, S. L.; Nussinov, R. *J. Mol. Biol.* **1997**, *265*, 68.
(19) Janin, J.; Miller, S.; Chothia, C. *J. Mol. Biol.* **1988**, *204*, 155.
(20) Jones, S.; Thornton, J. M. *Protein Eng.* **2000**, *13*, 77.
(21) Davies, D. R.; Cohen, G. H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 7.
(22) Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946.
(23) Norel, R.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *J. Mol. Biol.* **1995**, *252*, 263.
(24) Korn, A. P.; Burnett, R. M. *Proteins: Struct. Funct., Genet.* **1991**, *9*, 37.
(25) Horton, N.; Lewis, M. *Protein Sci.* **1992**, *1*, 169.
(26) McCoy, A. J.; Epa, V. C.; Colman, P. M. *J. Mol. Biol.* **1997**, *268*, 570.
(27) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
(28) Janin, J. *Structure* **1999**, *7*, R277.
(29) Otwinowski, Z.; Schevitz, R. W.; Zhang, R.-G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. *Nature* **1988**, *335*, 321.
(30) Brennan, R. G.; Matthews, B. G. *J. Biol. Chem.* **1989**, *264*, 1903.
(31) Joachimiak, A.; Haran, T. E.; Sigler, P. B. *EMBO J.* **1994**, *13*, 367.
(32) Smith, S. A.; Rajur, S. B.; McLaughlin, L. W. *Nat. Struct. Biol.* **1994**, *1*, 18.
(33) Dao-Pin, S.; Anderson, D. E.; Baase, W. A.; Dahlquist, F. W.; Matthews, B. W. *Biochemistry* **1991**, *30*, 11521.
(34) Stites, W. E.; Gittis, A. G.; Lattman, E. E.; Shortle, D. *J. Mol. Biol.* **1991**, *221*, 7.
(35) Yang, A.-S.; Honig, B. *Adv. Protein Chem.* **1995**, *46*, 27.
(36) Fersht, A. R. *Trends Biochem. Sci.* **1987**, *12*, 301.
(37) Wells, J. A. *Methods Enzymol.* **1991**, *202*, 390.
(38) Voet, D.; Voet, J. G. *Biochemistry*, 2nd ed.; Wiley: New York 1995.
(39) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1.
(40) Lichtarge, O.; Bourne, H. R.; Cohen, F. E. *J. Mol. Biol.* **1996**, *257*, 342.
(41) Hu, Z.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct. Funct. Genet.* **2000**, *39*, 331.
(42) Pazos, F.; Helmer-Citterich, M.; Ausiello, G.; Valencia, A. *J. Mol. Biol.* **1997**, *271*, 511.
(43) York, D. M.; Lee, T.-S.; Yang, W. *J. Am. Chem. Soc.* **1996**, *118*, 10940.
(44) Gogonea, V.; Merz, K. M. *J. Phys. Chem. A* **1999**, *103*, 5171.

(45) McCammon, J. A.; Harvey, S. C. *Dynamics of proteins and nucleic acids*; Cambridge University Press: Cambridge, U.K., 1987.
(46) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins. A theoretical perspective of dynamics, structure, and thermodynamics*; Wiley-Interscience: New York, 1988.
(47) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1991**, *29*, 992.
(48) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(49) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnik, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(50) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740.
(51) Eaton, W. A.; Munoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10.
(52) Zhou, H.-X.; Wlodek, S. T.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9280.
(53) Chillemi, G.; Falconi, M.; Amadei, A.; Zimatore, G.; Desideri, A.; DiNola, A. *Biophys. J.* **1997**, *73*, 1007.
(54) Wlodek, S. T.; Clark, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9513.
(55) Du, Q. H.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **2000**, *104*, 796.
(56) Roux, B.; Simonson, T. *Biophys. Chem.* **1999**, *78*, 1.
(57) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(58) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(59) Sham, Y. Y.; Muegge, I.; Warshel, A. *Biophys. J.* **1998**, *74*, 1744.
(60) Simonson, T. *Int. J. Quantum Chem.* **1999**, *73*, 45.
(61) Mertz, E. L.; Krishtalik, L. I. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2081.
(62) King, G.; Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366.
(63) Smith, P. E.; Brunne, R. M.; Mark, A. E.; van Gunsteren, W. F. *J. Phys. Chem.* **1993**, *97*, 2009.
(64) Simonson, T.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1996**, *118*, 8452.
(65) García-Moreno, E.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. A.; Spencer, D. S.; Stites, W. E. *Biophys. Chem.* **1997**, *64*, 211.
(66) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415.
(67) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 6142.
(68) Olson, M. A.; Reinke, L. T. *Proteins: Struct. Funct., Genet.* **2000**, *38*, 115.
(69) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.
(70) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, *186*, 815.
(71) Zhou, H.-X. *Biophys. J.* **1993**, *65*, 955.
(72) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671.
(73) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K. A.; Honig, B. *Proteins: Struct. Funct., Genet.* **1986**, *1*, 47.
(74) Nicholls, A.; Sharp, K. A.; Honig, B. *DelPhi*. Department of Biochemistry and Molecular Biophysics, Columbia University, New York, 1990.
(75) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. *Rev. Comput. Chem.* **1994**, *5*, 229.
(76) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, *91*, 57.
(77) Holst, M.; Baker, N.; Wang, F. *J. Comput. Chem.*, in press.
(78) Baker, N.; Holst, M.; Wang, F. *J. Comput. Chem.*, in press.
(79) Sharp, K.; Honig, B. *J. Phys. Chem.* **1990**, *94*, 7684.
(80) Hendsch, Z. S.; Tidor, B. *Protein Sci.* **1994**, *3*, 211.
(81) Misra, V. K.; Hecht, J. L.; Yang, A.-S.; Honig, B. *Biophys. J.* **1998**, *75*, 2262.
(82) Zacharias, M.; Luty, B. A.; Davis, M. E.; McCammon, J. A. *Biophys. J.* **1992**, *63*, 1280.
(83) Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245.
(84) Misra, V. K.; Hecht, J. L.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 264.
(85) Lounnas, V.; Wade, R. C. *Biochemistry* **1997**, *36*, 5402.
(86) Elcock, A. H. *J. Mol. Biol.* **1998**, *284*, 489.
(87) Xiao, L.; Honig, B. *J. Mol. Biol.* **1999**, *289*, 1435.
(88) Kumar, S.; Ma, B. Y.; Tsai, C. J.; Nussinov, R. *Proteins: Struct. Funct. Genet.* **2000**, *38*, 368.
(89) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1993**, *231*, 459.
(90) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819.
(91) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591.
(92) Gilson, M. K.; McCammon, J. A.; Madura, J. D. *J. Comput. Chem.* **1995**, *16*, 1081.
(93) Niedermeier, C.; Schulten, K. *Mol. Sim.* **1992**, *8*, 361.
(94) Elcock, A. H.; McCammon, J. A. *J. Am. Chem. Soc.* **1996**, *118*, 3787.
(95) David, L.; Luo, R.; Gilson, M. K. *J. Comput. Chem.* **2000**, *21*, 295.
(96) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868.
(97) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.
(98) Chothia, C. *Nature* **1974**, *248*, 338.
(99) Ashbaugh, H.; Kaler, E.; Paulaitis, M. *J. Am. Chem. Soc.* **1999**, *121*, 9243.
(100) Gallicchio, E.; Kubo, M. M.; Levy, R. M. *J. Phys. Chem. B* **2000**, *104*, 6271.
(101) Chothia, C.; Janin, J. *Nature* **1975**, *256*, 705.
(102) Sharp, K. A.; Nicholls, A.; Fine, R. F.; Honig, B. *Science* **1991**, *252*, 106.
(103) Friedman, R. A.; Honig, B. *Biophys. J.* **1995**, *69*, 1528.
(104) Ben-Naim, A. *Biopolymers* **1990**, *29*, 567.
(105) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199.
(106) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem.* **1995**, *47*, 307.
(107) Lazaridis, T.; Archontis, G.; Karplus, M. *Adv. Protein Chem.* **1995**, *47*, 231.
(108) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133.
(109) Massova, I.; Kollman, P. A. *Perspect. Drug Design* **2000**, *18*, 113.
(110) Janin, J. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 145.
(111) Sternberg, M. J. E.; Gabb, H. A.; Jackson, R. M. *Curr. Opin. Struct. Biol.* **1998**, *8*, 250.
(112) Python, M. *Monty Python and the Holy Grail*, Mandarin Paperbacks: London, 1993.
(113) Dill, K. A. *Protein Sci.* **1999**, *8*, 1166.
(114) Kim, S. H. *Curr. Opin. Struct. Biol.* **2000**, *10*, 380.
(115) Koonin, E. V.; Wolf, Y. I.; Aravind, L. *Adv. Protein Chem.* **2000**, *54*, 245.
(116) Ito, T.; Tashiro, K.; Muta, S.; Ozawa, R.; Chiba, T.; Nishizawa, M.; Tamamoto, K.; Kuhara, S.; Sakaki, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1143.
(117) Wodak, S. J.; Janin, J. *J. Mol. Biol.* **1978**, *124*, 323.
(118) Vakser, I. A.; Matar, O. G.; Lam, C. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8477.
(119) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesen, A. A.; Aflalo, C.; Wodak, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2195.
(120) Ten Eyck, L. F.; Mandell, J.; Roberts, V. A.; Pique, M. E. In *Proceedings of the 1995 ACM/IEEE Supercomputing Conference, San Diego, December 3−8, 1995*; Hayes, A., Simmons, M., Eds.; IEEE Computer Society Press: Los Alamitos, CA, 1995.
(121) Harrison, R. W.; Kourinov, I. V.; Andrews, L. C. *Protein Eng.* **1994**, *7*, 359.
(122) Weng, Z.; Vajda, S.; DeLisi, C. *Protein Sci.* **1996**, *5*, 614.
(123) Jackson, R. M.; Gabb, H. A.; Sternberg, M. J. E. *J. Mol. Biol.* **1998**, *276*, 265.
(124) Camacho, C. J.; Gatchell, D. W.; Kimura, S. R.; Vajda, S. *Proteins: Struct. Funct. Genet.* **2000**, *40*, 525.
(125) Moont, G.; Gabb, H. A.; Sternberg, M. J. E. *Proteins*: *Struct. Funct. Genet.* **1999**, *35*, 364.
(126) Totrov, M.; Abagyan, R. *Nature Struct. Biol.* **1994**, *1*, 259.
(127) Ponstingl, H.; Henrick, K.; Thornton, J. M. *Proteins: Struct. Funct., Genet.* **2000**, *41*, 47.
(128) Jackson, R. M. *Protein Sci.* **1999**, *8*, 603.
(129) Betts, M. J.; Sternberg, M. J. E. *Protein Eng.* **1999**, *12*, 271.
(130) Norel, R.; Petrey, D.; Wolfson, H. J.; Nussinov, R. *Proteins: Struct. Funct. Genet.* **1999**, *36*, 307.
(131) Covell, D. G.; Wallqvist, A. *J. Mol. Biol.* **1997**, *269*, 281.
(132) Strynadka, N. C. J.; Eisenstein, M.; Katchalski-Katzir, E.; Shoichet, B.; Kuntz, I.; Abagyan, R.; Totrov, M.; Janin, J.; Cherfils, J.; Zimmerman, F.; Olson, A.; Duncan, B.; Rao, M.; Jackson, R.; Sternberg, M.; James, M. N. G. *Nature Struct. Biol.* **1996**, *3*, 233.
(133) Shoichet, B. K.; Kuntz, I. D. *Chem. Biol.* **1996**, *3*, 151−156.
(134) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 133.
(135) Pelletier, H.; Kraut, J. *Science* **1992**, *258*, 1748.
(136) Srere, P. A.; Sherry, A. D.; Malloy, C. R.; Sumegi, B. In *Channelling in Intermediary Metabolism*; Agius, L., Sherratt, H. S. A., Eds.; Portland Press: London, 1997.
(137) McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945**, *13*, 276.
(138) Ree, F. H.; Hoover, W. G. *J. Chem. Phys.* **1967**, *46*, 4181.
(139) Minton, A. P. *Mol. Cell. Biochem.* **1983**, *55*, 119.
(140) Minton, A. P. *Biophys. Chem.* **1995**, *57*, 65.
(141) Neal, B. L.; Lenhoff, A. M. *AIChE J.* **1995**, *41*, 1010.
(142) Neal, B. L.; Asthagiri, D.; Lenhoff, A. M. *Biophys. J.* **1998**, *75*, 2469.

(143) Asthagiri, D.; Neal, B. L.; Lenhoff, A. M. *Biophys. Chem.* **1999**, *78*, 219.
(144) Chang, R. C.; Asthagiri, D.; Lenhoff, A. M. *Proteins: Struct. Funct., Genet.* **2000**, *41*, 123.
(145) Elcock, A. H.; McCammon, J. A. *Biophys. J.*, in press.
(146) Fersht, A. R. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; W. H. Freeman: New York, 1998.
(147) Taylor, P.; Radic, Z. *Annu. Rev. Pharm. Toxicol.* **1994**, *34*, 281.
(148) Ermak, D. L.; McCammon, J. A. *J. Chem. Phys.* **1978**, *69*, 1352.
(149) Oberholzer, M. R.; Lenhoff, A. M. *Langmuir* **1999**, *15*, 3905.
(150) Ravichandran, S.; Talbot, J. *Biophys. J.* **2000**, *78*, 110.
(151) Arora, A. K., Tata, B. V. R., Eds. *Ordering and phase transitions in charged colloids*; VCH Publishers: New York, 1996.
(152) Wade, R. C. *Biochem. Soc. Trans.* **1996**, *24*, 254.
(153) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Recognit.* **1999**, *12*, 226.
(154) Smoluchowski, M. V. *Z. Phys. Chem.* **1917**, *92*, 129.
(155) Creighton, T. E. *Proteins*; W. H. Freeman: New York, 1993.
(156) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *Biophys. J.* **2000**, *78*, 719.
(157) Sharp, K.; Fine, R.; Honig, B. *Science* **1987**, *236*, 1460.
(158) Getzoff, E. D.; Cabelli, D. E.; Fisher, C. L.; Parge, H. E.; Viezzoli, M. S.; Banci, L.; Hallewell, R. A. *Nature* **1992**, *358*, 347.
(159) Wade, R. C.; Luty, B. A.; Demchuk, E. D.; Madura, J. D.; Davis, M. E.; Briggs, J. M.; McCammon, J. A. *Nat. Struct. Biol.* **1994**, *1*, 65.
(160) Wade, R. C.; Gabdoulline, R. R.; Luty, B. A. *Proteins: Struct. Funct. Genet.* **1998**, *31*, 406.
(161) Radic, Z.; Kirchhoff, P.; Quinn, D. M.; McCammon, J. A.; Taylor, P. *J. Biol. Chem.* **1997**, *272*, 23265.
(162) Tara, S.; Elcock, A. H.; Kirchhoff, P. D.; Briggs, J. M.; Radic, Z.; Taylor, P.; McCammon, J. A. *Biopolymers* **1998**, *46*, 465.
(163) Gabdoulline, R. R.; Wade, R. C. *Methods: Companion to Methods Enzymol.* **1998**, *14*, 329.
(164) Northrup, S. H.; Boles, J. O.; Reynolds, J. C. L. *Science* **1988**, *241*, 67.
(165) Nambi, P.; Wierzbicki, A.; Allison, S. A. *J. Phys. Chem.* **1991**, *95*, 9595.
(166) Kozack, R. E.; Dmello, M. J.; Subramaniam, S. *Biophys. J.* **1995**, *68*, 807.
(167) Castro, G.; Boswell, C. A.; Northrup, S. H. *J. Biomol. Struct. Dyn.* **1998**, *16*, 413.
(168) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917.
(169) Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A. *J. Mol. Biol.* **1999**, *291*, 149.
(170) Sept, D.; Elcock, A. H.; McCammon, J. A. *J. Mol. Biol.* **1999**, *294*, 1181.
(171) Schreiber, G.; Fersht, A. R. *Nat. Struct. Biol.* **1996**, *3*, 427.
(172) Selzer, T.; Schreiber, G. *J. Mol. Biol.* **1999**, *287*, 409.
(173) Selzer, T.; Albeck, S.; Schreiber, G. *Nature Struct. Biol.* **2000**, *7*, 537.
(174) Vijayakumar, M.; Wong, K.-Y.; Schreiber, G.; Fersht, A. R.; Szabo, A.; Zhou, H.-X. *J. Mol. Biol.* **1998**, *278*, 1015.
(175) Jeltsch, A.; Wenz, C.; Pingoud, A. *EMBO J.* **1996**, *15*, 5104.
(176) Jeltsch, A.; Pingoud, A. *Biochemistry* **1998**, *37*, 2160.
(177) Adam, G.; Delbrück, M. in *Structural Chemistry and Molecular Biology*; Rich, A., Davidson, N., Eds.; Freeman, San Francisco, 1968; pp 198−215.
(178) von Hippel, P. H.; Berg, O. G. *J. Biol. Chem.* **1989**, *264*, 675.
(179) Camacho, C. J.; Kimura, S. R.; DeLisi, C.; Vajda, S. *Biophys. J.* **2000**, *78*, 1094.
(180) Northrup, S. H.; Erickson, H. P. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3338.
(181) Zhang, C.; Vasmatzis, G.; Cornette, J. L.; DeLisi, C. *J. Mol. Biol.* **1997**, *267*, 707.
(182) Leckband, D.; Sivasankar, S. *Colloids Surfaces B* **1999**, *14*, 83.
(183) Albright, R. A.; Mossing, M. C.; Matthews, B. W. *Protein Sci.* **1998**, *7*, 1485.
(184) Janin, J. *Nature Struct. Biol.* **1997**, *4*, 973.
(185) Henrick, K.; Thornton, J. M. *Trends Biochem. Sci.* **1998**, *23*, 358.
(186) Sharp, K. A. *Proteins* **1998**, *33*, 39.
(187) White, J. M. *Science* **1992**, *258*, 917.
(188) Laughlin, R. B.; Pines, D.; Schmalian, J.; Stojkovic, B. P.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 32.
(189) Bicout, D. J.; Field, M. J. *J. Phys. Chem.* **1996**, *100*, 2489.
(190) Ölveczky, B. P.; Verkman, A. S. *Biophys. J.* **1998**, *74*, 2722.