Mechanisms of Protein Assembly: Lessons from Minimalist Models

YAAKOV LEVY AND JOSÉ N. ONUCHIC*

Center for Theoretical Biological Physics, Department of Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093

Received May 26, 2005

ABSTRACT

Many cellular functions rely on interactions among proteins and between proteins and nucleic acids. The limited success of binding predictions may suggest that the physical and chemical principles of protein binding have to be revisited to correctly capture the essence of protein recognition. In this Account, we discuss the power of reduced models to study the physics of protein assembly. Since energetic frustration is sufficiently small, native topology-based models, which correspond to perfectly unfrustrated energy landscapes, have shown that binding mechanisms are robust and governed primarily by the protein's native topology. These models impressively capture many of the binding characteristics found in experiments and highlight the fundamental role of flexibility in binding. The essential role of solvent molecules and electrostatic interactions in binding is also discussed. Despite the success of the minimally frustrated models to describe the dynamics and mechanisms of binding, the actual degree of frustration has to be explored to quantify the capacity of a protein to bind specifically to other proteins. We have found that introducing mutations can significantly reduce specificity by introducing an additional binding mode. Deciphering and quantifying the key ingredients for biological self-assembly is invaluable to reading out genomic sequences and understanding cellular interaction networks.

Introduction

The life of cells is orchestrated by a network of chemical reactions involving numerous proteins and nucleic acids and the transport of those molecules between cellular compartments. The remarkable efficiency of organizing these processes to yield a cellular function presents a major theoretical puzzle given the large number of molecular species and the crowded environment they inhabit. In the recent years, we have come to understand the assembly of the individual actors in this drama thanks to many cooperative efforts between experiments and theory. We now understand the main principles of folding kinetics,[1,2] can often predict monomeric protein structure,[3] and can even design novel protein structures.[4] However, knowing everything about isolated monomeric proteins does not give a complete understanding of function. Function requires change of structure and specific recognition to form large assemblies. These processes must be governed by the information stored in their sequences and structures. Furthermore, biomacromolecules are flexible with a rich repertoire of movements on various length and time scales. These motions are essential to determine the ability of a protein to bind different ligands at the same or different binding sites.[5,6] Deciphering the molecular and structural origins of high specificity as well as the catalytic promiscuity and multitasking of proteins is prerequisite for a quantitative understanding of the complexity and multidimensionality in genomes. This cooperation of many proteins and nucleic acids, which is largely "wireless", is quite intricate. Understanding the principles of biomolecular assembly in quantitative detail constitutes the basis for the molecular theory of biological networks.

Theoretical and computational studies of protein binding have concentrated on analyzing the structural and chemical properties of interfaces[7,8] as well as predicting the structure of the formed complexes and their binding affinity.[9,10] Understanding the organization of proteins into large complexes is required to understand their function and irreversible aggregation. The challenge of predicting the complex formed between pairs of proteins has been addressed for several years by docking two proteins using various models, which range from reduced models[11,12] to atomistic ones[9] and include different flavors. Approaches to predict the structures of higher complexes, which are often defined as cellular machines, have been recently developed too. These approaches include, for example, combinatorial docking schemes[13] or fitting to cryoEM density maps at low resolution.[14] Some progress has been made in recent years in the performance of docking algorithms, yet their successes in predicting the structure of the protein complex are limited mainly to docking of the bound conformations of the complex subunits.

The inferiority of binding prediction to folding prediction is surprising because the conformational search required in binding processes of two folded proteins is smaller than that involved in protein folding. This shortcoming suggests that the physical and chemical principles of protein binding have to be revisited. The poor predictions of docking when using the conformations of the free subunits obviously indicates that protein flexibility is an important component in binding. Several docking approaches have introduced side-chain flexibility by using a rotamers library; however, it seems that backbone flexibility cannot be ignored.[15] It is likely, thus, that flexibility effects are still grossly underestimated as suggested from our recent association studies.[16−18] Solvent is also a critical component in protein association. While the protein cores are usually dry and contain a few water molecules, the interfaces in protein complexes are often very wet[19] (see Figure 1). Recently, it was found that a

Yaakov Levy received his Ph.D. from Tel-Aviv University, Israel, in 2002. He is currently a postdoctoral researcher in the groups of Profs. José Onuchic and Peter Wolynes at the University of California at San Diego. He will soon join the Department of Structural Biology at the Weizmann Institute as a Senior Lecturer. His current research interests focus on understanding the first principles of biological self-organization processes using theoretical and computational tools.

José Onuchic received his Ph.D. from Caltech in 1987. He joined the faculty of the University of California at San Diego in 1990, where he is currently Professor in Physics and co-director of the Center for Theoretical Biological Physics. His current research interests center on theoretical and computational methods for molecular biophysics and chemical reactions in condensed matter, with special emphasis in protein folding and electron transfer in biomolecules.

10.1021/ar040204a CCC: $33.50
Published on Web 10/15/2005

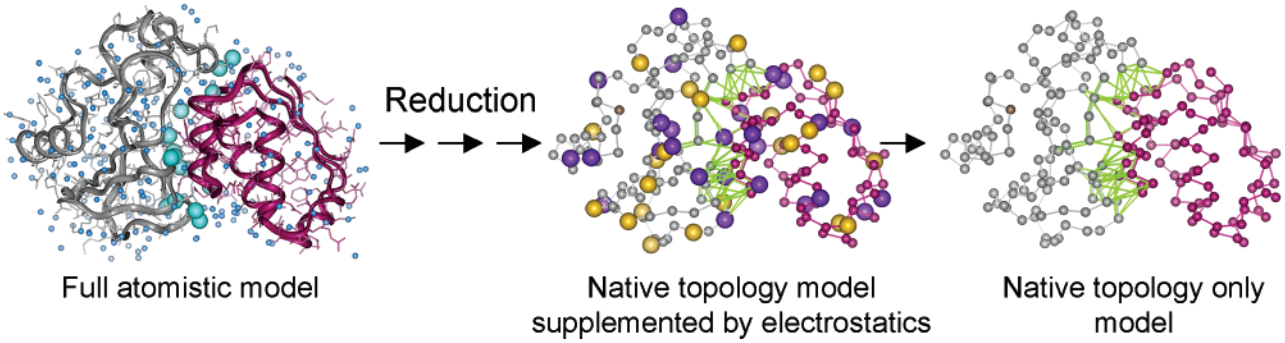

FIGURE 1. An illustration of the reduction procedure from full atomistic presentation to topological presentation at the residue level. The illustration is done on the complex between barnase and barstar (pdb entry 1BRS). The atomistic presentation includes 204 water molecules at the protein surface and 16 water molecules at the interface that mediate contacts between the two monomers. In the reduced models, each residue is represented by a single bead. In the native topology only model, all the beads are represented the same way and differ in their connectivity to other residues. To introduce long-ranged electrostatic forces, charges are associated with positively (arginine and lysine, designated by yellow spheres) and negatively (aspartic acid and glutamic acid, designated by purple spheres) charged residues. The green lines indicate interfacial interactions.

funneled potential for binding between proteins was obtained only upon solvation of assembly interfaces.[20] These observations provide a strong indication that water can be indispensable in protein assembly and undoubtedly in protein binding to DNA due to its highly charged surface.

Beyond structure prediction, we also have to understand the basic mechanisms of association processes in biology to understand protein function, binding specificity, and cross-talk in the cell. Quantifying the degree of flexibility in protein binding is necessary to understand the capability of a single protein to bind multiple unrelated ligands at the same binding site or at different sites. The need to understand the mechanism of binding and its main determinants is well illustrated by our study on the association pathway of dimeric HIV-1 protease.[21,22] These studies have indicated that the monomeric HIV-1 protease is relatively folded in its free form. The binding by association of prefolded monomers suggests a new way to inhibit the protease by designing an inhibitor that will bind to the monomer and thus prevent dimerization rather than designing an inhibitor that blocks the active site but will eventually become ineffective due to drug resistance.

In this Account, we will discuss the power of reduced models to address fundamental aspects of self-assembly of biological macromolecules. Reduced models are useful because of their computational efficiency in comparison to the heroic efforts required to study protein binding using all-atom simulations on a microsecond to second time scale. Moreover, reduced models with simplified presentations of biological macromolecules are especially attractive because they can be tailored to address a unique question. As such they allow one to model specific side chain information, electrostatic interactions, and solvent degrees of freedom, as well as nonadditive effects of a variety of levels of accuracy. In that respect, one may mention that the most significant progress in the theory of protein folding has been achieved by the development of simplified models that aimed to capture the physics of protein folding. In particular, we discuss models that are based on the complex native structure to study the mechanism of protein binding. We also discuss the need to develop other models to further investigate other aspects of biological self-organization processes.

Native Topology-Based Models of Protein Association

Native topology-based (Go) models for protein complexes include only the interactions that stabilize the native structure as determined by NMR or X-ray measurements. These models do not include nonnative interactions that are in conflict with the native structure. They, thus, correspond to an unfrustrated model with a perfectly funneled energy landscape where the native state is dominant and unique. Moreover, all the native interactions are modeled by a Lennard-Jones potential, which is short-ranged and very specific, and without discrimination based on the chemical property of the interactions. Figure 1 illustrates the reduction procedure from a full atomistic presentation to a topological presentation at the residue level, where each residue is identically represented or characterized by its charge. It has to be emphasized that reduced models can be built in various ways with the aim to serve the question at hand.

The native topology-based model has been recently applied in several studies to study the mechanism of protein association[16−18] and successfully reproduces the experimental classification of homodimers on whether monomer folding is prerequisite to monomer association. Obligatory homodimers that exhibit two-state thermodynamics are found to be formed by a coupled folding and binding reaction. Transient homodimers, which bind via a thermodynamic intermediate, are formed by association of already folded monomers. In general, we found that most of the gross and, surprisingly, many of the finer features of binding mechanisms can be obtained by Go simulations. The validity of the model to study protein binding is reflected by the good correlation obtained between the computational and experimental Φ values, which measure the degree of structure at the transition

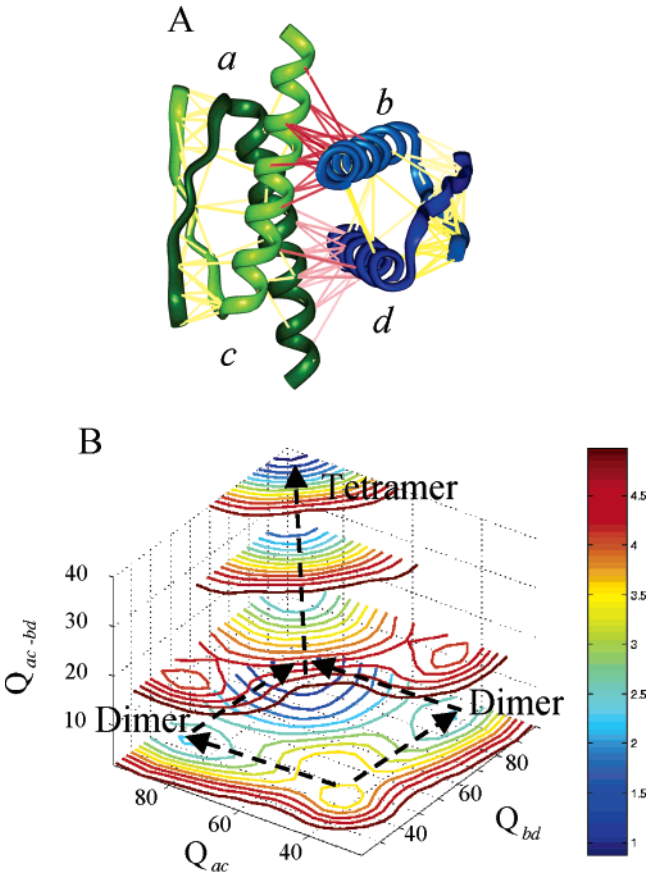

FIGURE 2. Association mechanism of p53tet. Panel A shows p53tet, which is a tetramer composed of four identical monomers (a—d). The interfacial contacts between the dimers composed of a and c as well as between b and d are shown by yellow lines. The interfacial contacts between the homodimers are shown by the red lines. Panel B shows the free energy landscape for the assembly along the reaction coordinates of formation of the dimers ac and bd and the tetramer. The free energy surface produced using the native topology-based model indicates that the tetramer is formed by dimerization of dimers, as found experimentally.

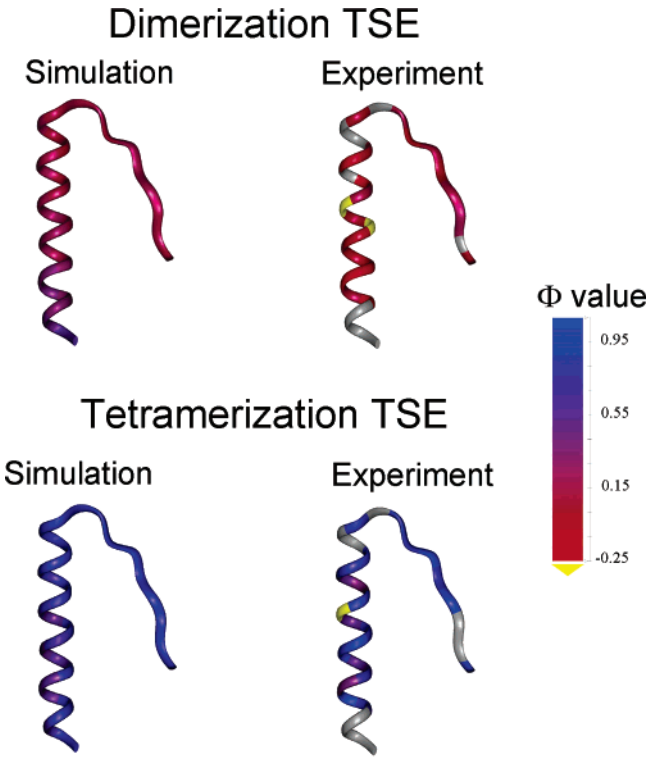

FIGURE 3. Comparison between the structure of the TSE from simulation and experiment. The two TSEs in the assembly of the tetramer (dimerization of monomers to form ac and bd and the dimerization of the dimer) were detected with the native topology-based model. The structure of the TSE is described by the residue Φ values. Residues that are colored red and blue correspond to Φ equal to 0 and 1, respectively. Residues that are colored yellow correspond to a Φ value below —0.5. A residue that is colored gray indicates unmeasured Φ value.

state ensemble (TSE) at the residue level. For Arc repressor and the tetramerization domain of p53 (p53tet), a direct comparison between the simulated and experimental Φ values is available and indicates that the simple Go model captures the nature of the transition state ensemble (TSE) reasonably well. For Arc repressor, there are detailed deviations between the simulated and experimental Φ values of particular residues (reflected by correlation coefficient of 0.31), but there is an agreement about the overall structure of the TSE. For p53tet, which was experimentally classified as a dimer of dimers, the native-centric model reproduced not only the association mechanism (Figure 2) but also the computational Φ values for the dimerization and tetramerization reactions in agreement with the experimental ones (Figure 3). We have to point out that recently an all-atom molecular dynamics study was done on the dimerization reaction of p53tet.[23] The Φ values for the binding TSE from that study, which includes nonnative interactions, display qualitatively similar results to those obtained from the native topology-based model.

The ability of native topology-based models to reproduce the features of binding mechanisms is significant and suggests that the binding TSE and binding mechanism can be obtained by the knowledge of the final complex's structure alone. A support for the role of topology in protein assembly can be found in a recent study that has shown that the folding rate of a two-state homoheptamer can be predicted on the basis of the topology of the native monomer.[24] We have recently found that protein complexes that are formed by the association of already folded subunits have structural and topological properties that are different from those with intrinsically unfolded subunits. More specifically, these two classes of complexes differ in the topological properties (i.e., connectivity of residues, average clustering coefficient, and mean shortest path length) of the monomers and the interfaces.[18]

Frustration in Protein Binding

The native topology-based model, which is an energetically unfrustrated model, impressively reproduces many features of protein associations, yet one may expect that adding nonnative contacts will improve the agreement between simulations and experiments. In folding studies, perturbing a native topology-based model toward a more realistic protein model by introducing nonnative interac-

tions results in an enhanced folding rate.[25] In binding reactions, nonnative interfacial interactions can assist specificity and increase binding rate. Alternatively, nonnative contacts can lead to trap states. Frustration in binding has to be quantified to understand the role of nonnative contacts and the degree of specificity in protein recognition by quantifying the effect of mutation at the binding site on the binding affinity. Moreover, frustration can result in several binding modes for a given pair of proteins where only one of them corresponds to the biological function. Frustration in binding is often observed in docking calculations that include many binding modes in addition to the correct one. Some of these binding modes can be physical; others can be false positive as a result of errors in the potential. To better understand how proteins are evolutionary designed for specific binding, we have to be able to quantify the properties of all other binding modes with the aim to decipher the code for correct binding.

Frustration at the cellular level may lead to cross-talk and promiscuous binding. A recent study has shown that a yeast peptide selectively binds to a single SH3 domain among the 27 SH3 domains in yeast; however, it promiscuously binds to 12 non-yeast SH3 domains.[26] Binding specificity in protein−protein and protein−DNA binding is fundamental in cellular regulation, and revealing its molecular origin is invaluable for understanding the cellular machinery and designing molecules that will bind on a network level.

Nature avoids frustration in protein interactions by evolutionarily designing sequences to fold efficiently and robustly to a unique structure and to bind specifically and selectively.[1] Some degree of frustration in biological systems is, however, unavoidable, and it has to be quantified.[25,27] The effect of frustration on binding specificity has been recently illustrated in the association modes of the repressor of primer (Rop) dimer.[28] We have shown that introducing mutations at the interface formed between the two identical monomers that composed the dimer can cause a large conformational change of the dimer. Mutating the interface in a symmetric fashion results in shifting the binding pathway of the two monomers. The variation in the structure of Rop mutants illustrates the capacity of frustration in affecting binding specificity. One has to note that the frustration exhibited in the Rop dimer was found in engineered mutants and the sequence of wild-type Rop, which is much less symmetric than its mutants, shows a distinct conformational specificity.

The role of minimally frustrated sequences of natural proteins in binding is also manifested by the prevalence of proteins to form domain-swapped oligomers. In this binding mode, a domain or a secondary structure element of a monomeric protein is exchanged with an identical region of another monomer, using the interactions that stabilize the monomeric state.[29] It was shown that from the topology of the monomer alone, the domain-swapped dimer of the protein can be correctly reproduced.[30] Some proteins, however, exhibit a larger degree of frustration because the interactions that define the monomeric state

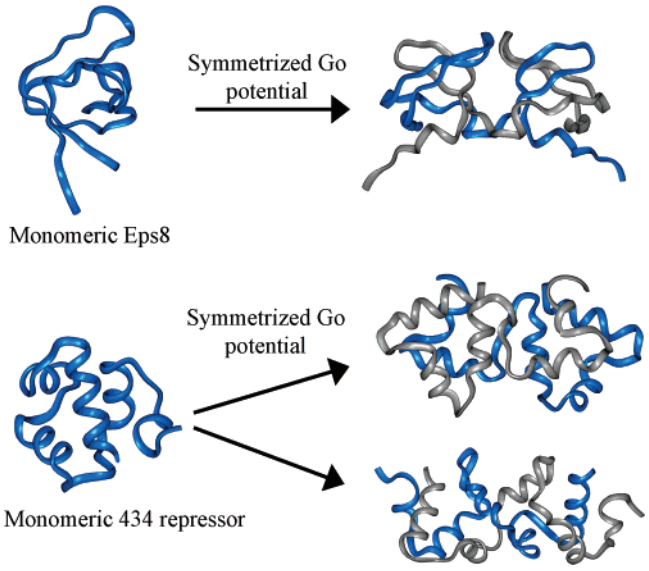

FIGURE 4. Prediction of the domain-swapped dimers of Eps8 and 434 repressor proteins with the symmetrized Go potential. The unique domain-swapped dimer obtained for Eps8 protein from the simulations is in agreement with the crystal structure. In addition to this structure, a few trap states of other swapped structures are found, however with relatively high free energy. For 434 repressor, however, two uniquely swapped structures with similar free energy were detected.

result in several domain-swapped structure.[30] Figure 4 shows the unique domain-swapped structure resulting when two monomers of Eps8 protein are simulated using the symmetrized Go potential that allows each intramonomeric contact to be formed intermolecularly as well. The unique swapped structure of Eps8 obtained with the symmetrized Go potential is the experimentally observed structure. Swapping of the monomers of 434 repressor does not result in a single structure but in at least two structures, suggesting that monomeric 434 repressor is frustrated for domain-swapping. A similar scenario was observed for the human prion protein.[30] However, introducing an intermolecular disulfide bond yielded the unique domain-swapped structure that was observed by X-ray crystallography. In other cases, where multiple swapped structures have been obtained, we conjecture that frustration in the monomer for binding via domain-swapping is needed to remain with a monomeric protein as the functional form of the protein.

The Need for Protein Flexibility in Protein Binding

Flexibility is pivotal in biological self-association processes. The role of flexibility in protein binding is acknowledged by Koshland's venerable "induced fit" mechanism, which expects conformational optimization upon binding. Structural rearrangement in binding is practically reflected by the very limited success of docking algorithms in predicting the complex structure when protein flexibility is not permitted or includes side-chain motion only.[15] Protein binding to DNA often involves, additionally to conformational change of the protein, bending of the DNA. Flex-

ibility and plasticity are central to protein folding when coupled to protein association or binding via the domain-swapping mechanism. Domain swapping requires at least partial unfolding of the monomers and the mutual interchange of symmetrically identical swapped regions. Large conformational changes upon binding should not be treated as rare exceptions. For example, intrinsically disordered proteins that form a perfectly ordered structure in the presence of the appropriate ligand constitute more than 30% of the genomes of several organisms, and the number of proteins observed in domain-swapped conformations constantly increases.

Our simulation studies indicate that protein flexibility is important also when the association is between folded proteins.[16−18,22] Formation of a symmetric protein complex can take place via an asymmetric pathway where one monomer is more folded than the other monomer. For λ cro repressor a more extreme case was observed where although its monomer is relatively stable on its own the association is between a folded monomer and an unfolded monomer that becomes folded in a later stage.[16,18] The association mechanism of λ cro repressor suggests that protein folding can be catalyzed in the presence of other monomers. The asymmetric binding pathway illustrates the demand for flexibility for an efficient binding that yield gradual recognition that is optimized upon complete binding. For λ cro repressor, the flexibility allows an unfolded monomer to bind to a folded monomer, which serves as a template, due to the simple geometrical and topological properties of the binding site. The minimalist model allows examination of the contribution of flexibility to binding by restriction of the protein dynamics. To assess the role of flexibility, the formation of p53tet was studied by simulating a dimerization reaction with partially flexible monomers and by simulating a tetramerization with relatively rigid dimers. A dimerization of less flexible monomers results in less efficient binding, emphasizing the crucial role of flexibility in the association of the monomers.[18] In a recent study on the binding mechanism of monomeric proteins to their binding sites on the DNA, we have found that the protein flexibility assists binding to DNA (unpublished results).

The Role of Electrostatic Interactions in Biological Recognition

Various surveys of the structures of protein complexes indicate the abundance of charged and polar residues in protein−protein interfaces[8,31] reflecting the possibly crucial role of electrostatic interactions in binding processes. While the contributions of nonpolar interactions to protein complexes are acknowledged, the roles of electrostatic interactions have been controversial. The energetic contribution of electrostatic interactions can be viewed as arising from a balance between the gain of pairwise Coulomb interactions and the loss of interactions with the solvent. In a number of studies electrostatic interactions have, on net, been found to oppose complex formation (the driving force came mainly from nonpolar interactions) and in other cases electrostatics provides a significant driving force favoring binding.[32,33] In addition, the effects of electrostatic interactions on specificity and binding rate were questioned. Several studies have reported that electrostatic forces can accelerate recognition by the long-range electrostatic attraction that controls the formation of the encounter complex, which is further stabilized by hydrophobic contacts and hydrogen bonds to form the high-affinity complex.[32,34−36] The complex role of electrostatic interactions in binding is reflected by salt bridges of coiled-coil dimer that stabilize but do not accelerate association.[37]

To explore the role of electrostatics in binding recognition, we supplemented the native topology based model by pairwise Coulomb interactions that are not specific. Since we are interested in studying the effects of long-range electrostatic interactions on the binding mechanism, repulsive and attractive electrostatic forces are allowed intermolecularly but not intramolecularly. We used as a system model the complex formed between barnase and its natural inhibitor, barstar, which exhibits high electrostatic complementarity and a tight interface (Figure 1). The binding between barnase and barstar has attracted a great deal of attention among protein−protein interactions, both experimentally and theoretically, particularly to explore the role of electrostatics on binding. While electrostatics was suggested to speed up association of barnase and barstar,[32] there are contradictory reports on its effect on the binding energetics.[38]

Our reduced model, which includes topology and electrostatics, is tailored to study the interplay between protein flexibility and electrostatic forces. The space of the protein binding Hamiltonian is explored by scanning the dielectric constant in the range of 25−150 and also when electrostatics was turned off (i.e., infinite dielectric constant). The specific heat profiles during the formation of the complex between barnase and barstar at different strengths of electrostatic interaction are shown in Figure 5A. These profiles include two peaks associated with the folding of the two monomeric proteins. Barnase has a higher folding temperature and broader C_v curve, reflecting its greater stability than barstar and its folding via an intermediate.[39] Introducing electrostatics results in a slight destabilization of the two monomers, presumably as a result of repulsive interactions; however, an enhanced effect for stronger electrostatics is not observed. One may note that each monomer is more stable in the environment of the dimer than when it is isolated for any strength of electrostatic forces. This phenomenon, which was observed for other homodimeric proteins,[22] can be attributed to a like-crowding effect that stabilizes proteins.[40]

The folding of barstar was found to be coupled to its binding to a folded barnase protein. This coupled folding and binding is observed for all the studied strengths of electrostatic interactions (Figure 6), indicating that barstar is optimized by evolution for function (i.e., tight and efficient binding) but not for high stability. The unique properties of the interface in the complex of barnase and barstar were recently illustrated by cluster analysis of

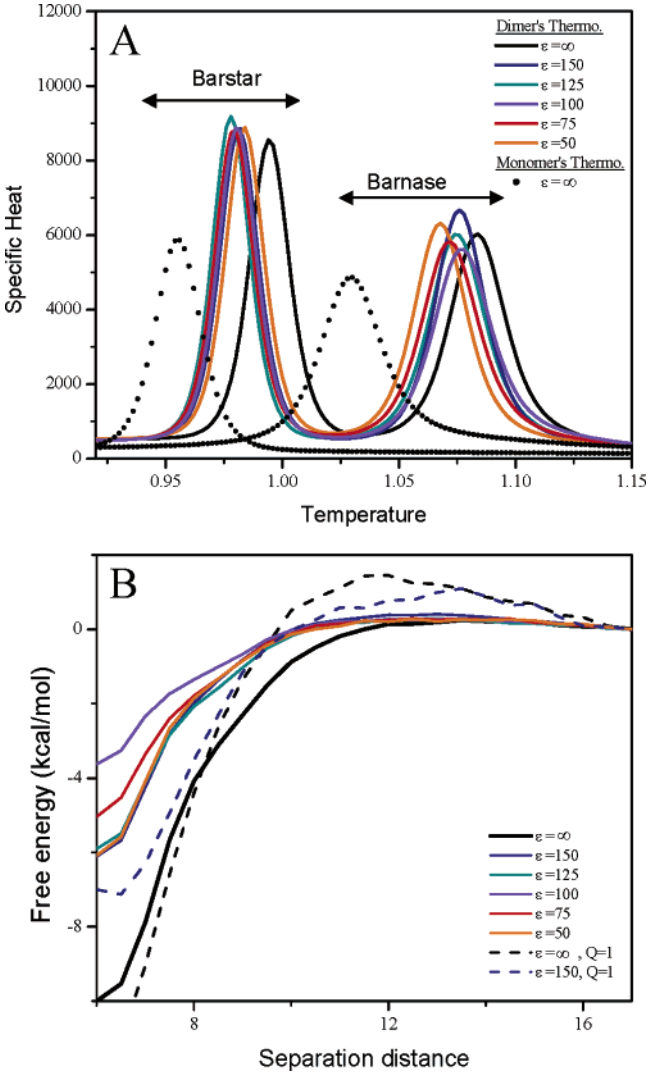

FIGURE 5. Association mechanism of the complex between barnase and barstar. Panel A shows the specific heat as a function of temperature for different values of dielectric constant. The dotted lines correspond to the free barnase and barstar. The two peaks correspond to folding of the two subunits. Panel B shows the free energy as a function of the separation distance during the dimerization reaction.

protein−protein interactions.[41] The possible role of flexibility in the recognition between barstar and barnase is supported experimentally by its extremely lower denaturation enthalpy in comparison to other small globular proteins.[42,43] Although free barstar is structured in solution and has similar structure to that in its complexed state,[44] it has a poorly packed hydrophobic interior and thus high conformational flexibility in the native state. The highly dynamic hydrophobic core and marginal stability of barstar are supported by NMR data and atomistic molecular dynamic simulations.[45] Its intrinsic flexibility is also reflected by the observation that satisfactory docking between barstar and barnase was achieved only when backbone flexibility was added to side-chain flexibility.[15] When electrostatic forces are included binding can occur also by binding between folded subunits; however, the coupled folding/binding mode is still dominant, reflecting the intrinsic flexibility of barstar and its role in binding.

Electrostatic interactions can destabilize the unbound state by repulsion between the two subunits and their folding prior to binding. The effect of electrostatic interaction is seen in Figure 5B when the free energy of the association is plotted as a function of the distance between the center of mass of the two subunits. For all values of simulated dielectric constants, a decrease in the free energy is observed when the two subunits are 3−6 Å further apart relative to the optimal separation distance between them in the crystal structure. A higher degree of flexibility is involved in the binding when all interfacial contacts are modeled by specific short-range interactions than when electrostatic long-range interactions are included. The crucial role of flexibility in formation of the barnase−barstar complex is reflected by the barrier obtained when binding is constrained to occur between rigid proteins (with or without electrostatics).

The Role of Water in Binding

The solvent is widely appreciated to be important in governing protein folding and binding. Yet, its exact effects and roles are not completely understood. Recently, it was reported that adding water to a Hamiltonian for structure prediction improved the predicted structures,[46] suggesting that water is an integral part of the structure. A dominant role of water in binding is obviously expected in protein binding, especially when the binding is between two relatively folded proteins with hydrophilic surfaces.[19] The abundance of water molecules in protein interfaces as detected by X-ray crystallography, although underestimated, serves as simple evidence for their structural role and as evidence that they are indispensable for protein recognition (see Figure 1). For some complexes, the enthalpy gain from water-mediated contacts is greater than the entropic cost that must be paid for immobilizing interfacial water. In a recent simulation study, it was proposed that a binding reaction can be driven by entropy increase as a result of bound water in a relatively large pocket.[47] A structural or functional role of water is even more obvious in binding of proteins to DNA interfaces, which are highly solvated.

Our understanding of the role of water solvation in protein folding has improved; however, the limited successes of implicit solvent models to accurately represent protein stability and dynamics suggest that the physics of the interaction between biomolecules and the solvent is not completely captured.[48,49] Reduced models with simple representations of protein solvation have already started to shed light on the physics of protein−water interactions[46,50] and are essential to understand fundamental questions in binding and recognition. The solvent molecules are likely to govern binding kinetics and stability. Although the binding mechanisms and transition states are correctly characterized by topologically based models that can effectively take into account structural water molecules but not dynamic water molecules,[18,19] introducing solvent effects to binding models is required to study desolvation effects, which can dominate the first

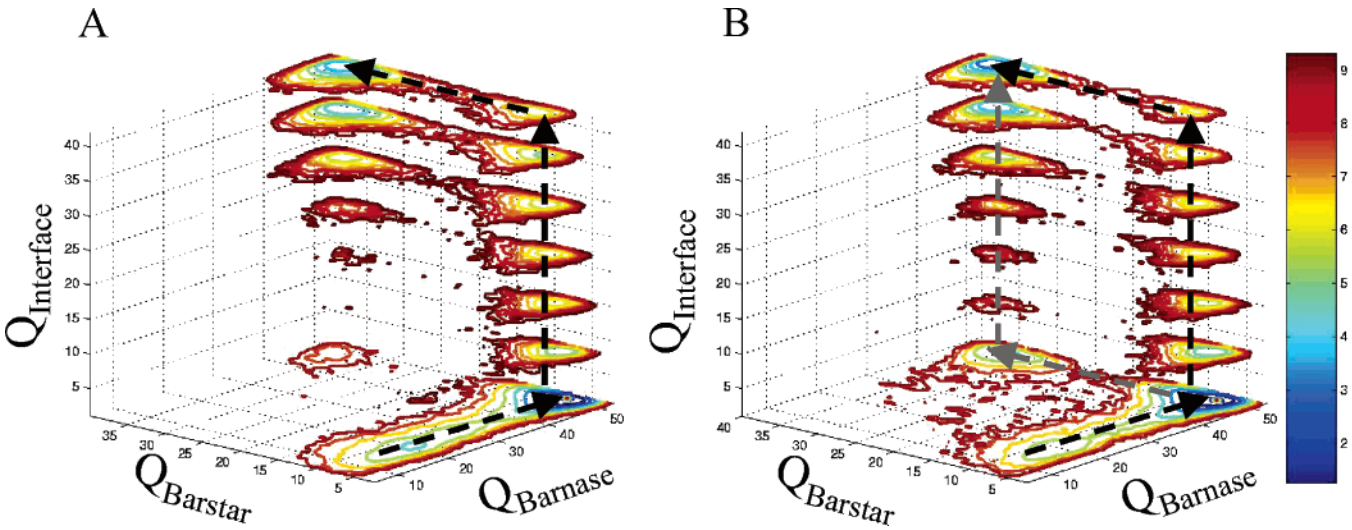

FIGURE 6. Four dimensional energy landscapes for the association of barnase and barstar with no electrostatic forces (A) and with dielectric constant equals to 75 (B). The landscapes are plotted along the reaction coordinates of folding of barnase and barstar and formation of the interface (the green lines in Figure 1).

stage of binding. We had conjectured that our simulations of antibody−antigen complex using the topology-based model poorly reproduced the binding TSE due to lack of water molecules in our model.[18] The abundance of water in mediating contacts in other forms of the complex managed to explain the discrepancy between the experimental and computational characterization of the binding transition state.[34] Solvent molecules, thus, can assist the initial association to form the encounter complex or, alternatively, the main binding transition state, which will be squeezed out at a later stage and result in a dry interface, which is stabilized by shape complementarity.

Conclusions and Perspectives

To understand the cellular network organization and how the genome is read out, we have to understand binding and recognition processes at the molecular level. Since all-atom molecular dynamics simulations of protein binding and protein−DNA recognition at physiological conditions will remain impractical in the foreseeable future, reduced models have to be developed to decipher the principles of self-organization in the cell from a physical viewpoint. Several aspects of binding have been discussed in this paper, and their investigation is needed to capture cellular communication. Beyond improving our ability to predict the complex formed between interacting proteins, one has to be able to predict their binding affinity and association rate. While the structural aspects of protein complexes have been addressed, and there are some successes in predicting complex structures, our understanding of the physics of macromolecular assembly is limited. We discussed in this Account the power of native topology-based models to study the mechanisms of biological binding. The models propose the importance of flexibility, as well as electrostatic and water-mediated interactions, for the assembly of some proteins, yet the magnitude of these effects has to be quantified in the future.

Beyond that, in the genomic era, we have to study protein−protein interactions in the context of the cell. The crowded cellular environment may affect the complex stability and its formation pathway. Moreover, the coexistence of many proteins and nucleic acids requires the understanding of specificity in recognition and how proteins are evolutionary designed to avoid cross-reactivity. The interplays between binding specificity, affinity, promiscuity, and the protein plasticity in both protein−protein and protein−DNA interactions will ultimately give meaning to raw genomics.

We would like to thank Samuel Cho for editing the manuscript and for providing the data for Figure 4. This work has been funded by the NSF sponsored Center for Theoretical Biological Physics (Grants PHY-0216576 and -0225630) with additional support from MCB-0084797.

References

(1) Onuchic, J. N.; Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(2) Kubelka, J.; Hofrichter, J.; Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.
(3) Venclovas, C.; Zemla, A.; Fidelis, K.; Moult, J. Assessment of progress over the CASP experiments. *Proteins* **2003**, *53* (6), 585−595.
(4) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003**, *302*, 1364−1368.
(5) Jeffery, C. J. Moonlighting proteins. *Trends Biochem. Sci.* **1999**, *24*, 8−11.
(6) James, L. C.; Tawfik, D. S. Conformational diversity and protein evolution-a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **2003**, *28*, 361−368.
(7) Jones, S.; Thornton, J. M. Principles of protein−protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13−20.
(8) Conte, L. L.; Chotia, C.; Janin, J. The atomic structure of protein−protein recognition sites. *J. Mol. Biol.* **1999**, *285*, 2177−2198.
(9) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47*, 409−443.
(10) Wodak, S. J.; Mendez, R. Prediction of protein−protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* **2004**, *14*, 242−249.
(11) Vakser, I. A. Protein docking for low-resolution structures. *Protein Eng.* **1995**, *8*, 371−377.

(12) Lu, H.; Lu, L.; Skolnick, J. Development of unified statistical potentials describing protein−protein interactions. *Biophys. J.* **2003**, *84*, 1895−1901.
(13) Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. Prediction of Multimolecular Assemblies by Multiple Docking. *J. Mol. Biol.* **2005**, in press.
(14) Russell, R. B.; Alber, F.; Aloy, P.; Davis, F. P.; Korkin, D.; Pichaud, M.; Topf, M.; Sali, A. A structural perspective on protein−protein interactions. *Curr. Opin. Struct. Biol.* **2004**, *14*, 313−324.
(15) Ehrlich, L. P.; Nilges, M.; Wade, R. C. The impact of protein flexibility on protein−protein docking. *Proteins* **2005**, *58*, 126−133.
(16) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. Protein topology determines binding mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 511−516.
(17) Levy, Y.; Papoian, G. A.; Onuchic, J.; Wolynes, P. G. Energy landscape of protein dimers. *Isr. J. Chem.* **2004**, *44*, 281−297.
(18) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J. Mol. Biol.* **2005**, *346*, 1121−1145.
(19) Levy, Y.; Onuchic, J. N. Water and proteins: A love-hate relationship. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3325−3326.
(20) Papoian, G. A.; Ulander, J.; Wolynes, P. G. Role of water mediated interactions in protein−protein recognition landscapes. *J. Am. Chem. Soc.* **2003**, *125*, 9170−9178.
(21) Levy, Y.; Caflisch, A. The flexibility of monomeric and dimeric HIV-1 PR. *J. Phys. Chem. B* **2003**, *107*, 3068−3079.
(22) Levy, Y.; Caflisch, A.; Onuchic, J. N.; Wolynes, P. G. The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations. *J. Mol. Biol.* **2004**, *340*, 67−79.
(23) Chong, L. T.; Snow, C. D.; Rhee, Y. M.; Pande, V. S. Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations. *J. Mol. Biol.* **2005**, *345*, 869−878.
(24) Bascos, N.; Guidry, J.; Wittung-Stafshede, P. Monomer topology defines folding speed of heptamer. *Protein Sci.* **2004**, *13*, 1317−1321.
(25) Clementi, C.; Plotkin, S. S. The effects of nonnative interactions on protein folding rates: theory and simulation. *Protein Sci.* **2004**, *13*, 1750−1766.
(26) Zarrinpar, A.; Park, S. H.; Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **2003**, *426*, 676−680.
(27) Wang, J.; Verkhivker, G. M. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys. Rev. Lett.* **2003**, *90*, No. 188101.
(28) Levy, Y.; Cho, S. S.; Shen, T.; Onuchic, J. N.; Wolynes, P. G. Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2373−2378.
(29) Liu, Y.; Eisenberg, D. 3D domain swapping: As domains continue to swap *Protein Sci.* **2002**, *11*, 1285−1299.
(30) Cho, S. S.; Levy, Y.; Onuchic, J. N.; Wolynes, P. G. Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation. *Physical Biology* **2005**, *2*, S44−S55.
(31) Xu, D.; Lin, S.; Nussinov, R. Protein binding versus protein folding: the role of hydrophilic bridges in protein association. *J. Mol. Biol.* **1997**, *1997*, 68−84.
(32) Schreiber, G.; Fersht, A. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* **1996**, *3*, 427−431.
(33) Sheinerman, F. B.; Honig, B. On the role of electrostatic interactions in the design of protein−protein interactions. *J. Mol. Biol.* **2002**, *318*, 161−177.
(34) Schreiber, G. Kinetic studies of protein−protein interactions. *Curr. Opin. Struct. Biol.* **2002**, *12*, 41−47.
(35) Elcock, A.; Gabdoulline, R.; Wade, R.; McCammon, J. Computer Simulation of Protein−Protein Association Kinetics: Acetylcholinesterase-Fasciculin. *J. Mol. Biol.* **1999**, *291*, 149−162.
(36) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic aspects of protein−protein interactions. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153−159.
(37) Ibarra-Molero, B.; Zitzewitz, J. A.; Matthews, C. R. Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide, GCN4-p1 *J. Mol. Biol.* **2004**, *336*, 989−996.
(38) Wang, T.; Tomic, S.; Gabdoulline, R. R.; Wade, R. C. How optimal are the binding energetics of barnase and barstar? *Biophys. J.* **2004**, *87*, 1618−1630.
(39) Fersht, A. R. A kinetically significant intermediate in the folding of barnase. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14121−14126.
(40) Klimov, D. K.; Newfield, D.; Thirumalai, D. Simulations of beta-hairpin folding confined to spherical pores using distributed computing. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 8019−8024.
(41) Reichmann, D.; Rahat, O.; Albeck, S.; Meged, R.; Dym, O.; Schreiber, G. The modular architecture of protein−protein binding interfaces *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 57−62.
(42) Agashe, V. R.; Shastry, M. C.; Udgaonkar, J. B. Initial hydrophobic collapse in the folding of barstar. *Nature* **1995**, *377*, 754−757.
(43) Wintrode, P. L.; Griko, Y. V.; Privalov, P. L. Structural energetics of barstar studied by differential scanning microcalorimetry. *Protein Sci.* **1995**, *4*, 1528−1534.
(44) Ratnaparkhi, G. S.; Ramachandran, S.; Udgaonkar, J. B.; Varadarajan, R. Discrepancies between the NMR and X-ray structures of uncomplexed barstar: analysis suggests that packing densities of protein structures determined by NMR are unreliable. *Biochemistry* **1998**, *37*, 6958−6966.
(45) Wong, K. B.; Daggett, V. Barstar has a highly dynamic hydrophobic core: evidence from molecular dynamics simulations and nuclear magnetic resonance relaxation data. *Biochemistry* **1998**, *37*, 11182−11192.
(46) Papoian, G. A.; Ulander, J.; Eastwood, M. E.; Wolynes, P. G. Water in protein structure prediction. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3352−3357.
(47) Petrone, P. M.; Garcia, A. E. MHC-peptide binding is assisted by bound water molecules. *J. Mol. Biol.* **2004**, *338*, 419−435.
(48) Garcia, A. E.; Hummer, G. Water penetration and escape in proteins *Proteins* **2000**, *38*, 261−272.
(49) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. Simulations of the role of water in the protein-folding mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6456−6461.
(50) Cheung, M. S.; Garcia, A. E.; Onuchic, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 685−690.

AR040204A