Energy Landscape Analysis of Protein Dimers

YAAKOV LEVY,[a,b,*] GAREGIN A. PAPOIAN,[b] JOSÉ N. ONUCHIC,[a] AND PETER G. WOLYNES[a,b]
[a]Center for Theoretical Biological Physics and Department of Physics, [b]Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

(Received 7 August 2003 and in revised form 23 August 2003)

Abstract. Many cellular functions are carried out by proteins that are bound together in multiprotein complexes. The binding between two highly flexible proteins to form homodimers is studied here using energy landscape theory and simulations based on a perfectly funneled energy landscape. With the aim to survey the range of binding mechanisms, two sets of homodimers were selected based on the experimental knowledge of whether stable monomers are needed for binding to take place. We find that the binding mechanism can be predicted based on the structure of the complex subunits alone. On average, the theory predicts a lower stability for subunits that are less compact and less hydrophobic, indicating, in agreement with their experimental classification, that their folding will be coupled to their binding. On the other hand, when a monomeric intermediate is experimentally found, the predicted stability of the monomers is comparable to that of known folded proteins. Furthermore, when dimerization is coupled to monomer folding, the interface is more hydrophobic.

INTRODUCTION

Most proteins function by interacting with partners that are small molecules or, very often, other biological macromolecules: proteins, nucleic acids, or polysaccharides. Protein–protein and protein–nucleic acid interactions are ubiquitous and fundamental to many cellular processes. Recent studies of protein complexes in yeast have demonstrated that most proteins exist in the cell as parts of multicomponent assemblies.[1,2] Furthermore, most of these complexes have components in common with at least one other multiprotein complex, reflecting a high-order network of interacting protein complexes. To predict such potential interactions on the proteomic scale,[3,4] there is a need for simple theoretical and computational approaches to help understand the dynamics and specificity of protein recognition and assembly. Such understanding may lead to the ability to design more stable complexes, which can act as "network" drugs. We hope that deciphering protein–macromolecule interactions will help reveal the large-scale patterns of protein networks responsible for higher-level properties of organisms, such as robustness and error correction.[5] Understanding protein–protein interactions may also help us to understand pathogenic irreversible aggregations of proteins that lie at the root of many diseases.[6]

Various mechanisms have been envisioned for protein binding. They vary in the role played by the flexibility of the complex components during their assembly. The paradigm that protein function is strictly related to the three-dimensional structure of the folded state has led many to believe binding occurs by the association of already folded proteins. The earliest proposed mechanism of biomolecular recognition describes association of the folded proteins as rigid body docking (the "lock and key" mechanism[7]). The flexibility of the protein in its folded state was later invoked in mechanisms that

We are happy to dedicate this paper to Joshua Jortner, whose inspiring contributions to chemical physics range from nuclear to biological phenomena. His seminal contributions in the field of cluster physics, we hope to have shown here, have their analogue in biomolecular problems.

*Author to whom correspondence should be addressed. E-mail: klevy@physics.ucsd.edu

suggest there is a selection of the correct conformer for binding (the "conformational selection"[8,9]), or that interactions formed initially in the encounter complex are later optimized in slow conformational transitions (the "induced fit"[10] mechanisms). Still larger flexibility is envisioned in the domain-swapping mechanism of protein binding.[11,12] In this mechanism, proteins assemble by exchanging a secondary structure element or an entire globular domain with the symmetrically identical part of the other subunit, an interaction consistent with the principle of minimal frustration.[13–16] Proteins that form domain-swapped oligomers have two well-defined native states: a monomeric and an intertwined structure. The mechanism of assembly by domain-swapping and its biological significance in both functional protein assemblies and in pathologic aggregates are therefore of much interest.

Protein flexibility is even more important when the monomer folding is directly coupled to binding events.[17–19] A good deal of evidence indicates that in vivo there are disordered proteins within the cell that fold only upon binding.[20] A sequence-based bioinformatics approach has predicted that more than 30% of the genome of 29 eukaryotes have proteins with disordered regions of 40 or more consecutive residues.[21] An energy landscape survey of a large database of protein complexes has suggested that ~15% of monomers may not fold in the absence of partner proteins.[22] Has this behavior been selected, or is it the easiest solution for evolution? While evolutionary drift is certainly part of the answer,[16] several advantages have been suggested for the use of disordered proteins that only fold upon reaching their targets. One selective advantage is that natively unfolded proteins are more adaptive, giving them the capability to bind to several different targets,[23,24] overcome steric clashes, and thus to achieve high specificity with low affinity.[25] Another advantage of being unfolded is the capability to form complexes with large interfaces, which may therefore contain more information. For a protein to be stable as a monomer while having extensive interfaces, the size of the protein needs to be more than twice as large as one with a small interface, resulting in increased cellular crowding.[26] Accordingly, disordered proteins provide a simple solution to having large intermolecular interfaces, while maintaining a small genome. A kinetic advantage for being initially unfolded before binding has been postulated through the fly-casting mechanism.[27] A partially structured or unstructured protein has a greater capture radius than a folded protein with its limited flexibility for a specific binding site, thereby enhancing the speed of association.

Folding and binding share similar characteristics and perhaps may be viewed as analogous processes where nonbonded interactions are formed intra- or intermolecularly. When the binding process is coupled to monomer folding, the search problem is similar to that of protein folding. When binding occurs between already folded subunits, the search space is smaller than for folding, but still large enough that predicting the structure of a complex formed between two interacting proteins is a challenge.[28–30] The search involved in binding is even more extensive when we also take into account all the possible complexes a protein can form in a cell with inappropriate partners. In addition, folding and binding are similar because both are driven by forces strongly modulated by the solvent environment.

We have recently carried out a simulation study of the formation of various homodimers. This study provided a strong indication that, similarly to protein folding, binding processes are guided by a funneled energy landscape.[31] In that survey, the association mechanisms of more than 10 homodimers were explored with a simplified model that includes only those contacts that exist in the native structure. The simulation model, accordingly, corresponds to a perfectly funneled landscape. The agreement between the mechanisms predicted by the simulations using this energetically unfrustrated model and the experimental classification of the various dimers as to whether there are intermediates during binding indicates indeed that binding processes are likely to be funneled. In a different study, an analytical energy landscape theory of coupling between binding and folding was developed, which showed that for some protein complexes, the initiation of folding is dependent on the formation of native binding contacts.[22] The funneled landscape for folding and binding processes suggests that proteins are evolutionarily designed to follow the principle of minimal frustration,[32,33] which results in a faster search through the many alternatives in the cell and considerable robustness of binding to mutations. The funneled shape toward the native binding state guarantees that binding will be stable against environmental and evolutionary fluctuations. The funnel concept was previously used to explain different binding mechanisms,[8,22,34–36] enzyme pathway and allostery,[37,38] binding selectivity and specificity,[39] and the role of water-mediated interactions in enhancing recognition in binding.[40]

In the present work, we combine analytical and simulation approaches to examine the properties of dimeric proteins that either fold only upon binding (2-state binding) or that fold via a monomeric intermediate (3-state binding). These are studied by (i) structural evaluation, (ii) thermodynamic analysis using the energy landscape theory, and (iii) simulations of simple models that

correspond to a funneled energy landscape (the so-called Gō model[41]). The binding kinetics is analyzed in greater detail for three homodimers that follow different binding mechanisms. The selected homodimers are Arc and lambda repressors. These function as transcription regulators by binding to a single operator site. The other protein singled out for examination is bovine seminal ribonuclease (BS-RNase). This molecule enzymatically cleaves RNA. While no monomeric Arc repressor was detected during the binding of two identical unfolded chains (2-state mechanism),[42] a folded monomer was observed for lambda repressor[43] that constitutes an obligatory intermediate before binding occurs (3-state mechanism); and for BS-RNase, two native monomeric states were detected (domain-swapping mechanism).[44] The thermodynamics and kinetics of the three different binding scenarios are discussed.

MODELS AND METHODS

Energy Landscape Analysis of Monomer Stability

To analyze the trends in the thermodynamic stabilities of monomers that comprise the structures of homodimers studied in the current work, we have computed the free energies of folding within the Minimally Frustrated Random Energy Model (MFREM) of protein folding.[45] In addition, in the spirit of the calibration procedure suggested earlier,[22] we have used the same model to compute the thermodynamic stabilities of 210 monomeric proteins that are known to fold independently.[46]

According to the MFREM theory of protein folding, the free energy change during the folding process is described as

$$\delta F = \delta E - \left(-S^0 T - \frac{\Delta \varepsilon^2}{2k_B T} \right) \qquad (1)$$

where δE indicates the energy gap between the average energies of the native states and the ones in the denaturated ensemble. In this model, larger δE leads to a deeper funnel, i.e., the native states become more favored thermodynamically. The denaturated ensemble, on the other hand, is thermodynamically stabilized when either the configurational entropy of the disordered phase (S^0) or the ruggedness of its energy landscape ($\Delta\varepsilon^2$) becomes large. Larger energetic ruggedness also leads to kinetic trapping, an effect not explicitly explored in this paper.

The parameters that enter eq 1 were chosen according to the procedure described earlier.[22] In particular, for a given pairwise potential (described below), 10 000 decoy sequences were generated by permutating the original sequence. We then compute the mean energy and the energy variance of the resulting model of the disordered ensemble. The variances were scaled down by a uniform constant to keep the folding temperature above the glass transition temperature. This scaling is required because the minimalist side chain model is by no means perfect. The configurational entropy per residue was chosen as $3k_B$.[47] The reference folding temperature for the calibration set of 210 proteins was set at 340 K, while all free energy calculations were carried out at 300 K.

Residue-based pair potentials, while not very accurate, are often used to model the complex potential energy surfaces of protein configurations. In this coarse-grained description, two residues are considered in contact if their C_β atoms (C_α for Gly) are within some distance cutoff (6.5 Å) of each other. In a previous paper,[40] a method based on energy landscape theory was developed to derive residue–residue pairwise potentials suitable for folding and binding. In this work, we have used a similar, but purely folding pairwise potential developed in the spirit of the previous work,[40] based on a random selection of 150 training proteins from the 210-protein monomer database.[46]

Structural Analysis of Dimers

With the aim to elucidate the structural and thermodynamic properties that govern two- and three-state binding mechanisms, two sets of homodimers that obey these mechanisms were selected. Besides ensuring that many examples from the experimental classification of the dimer association mechanism were included, the dimers were selected to span a range of topology, secondary structure content, and interface geometry (Table 1). The dimer structure may be most crudely described by the number of interactions in the folded state (the total number of native contacts). An interaction between a pair of residues (i, j) exists if at least a couple of atoms belonging to residues i and j in the native structure are considered to be in contact according to the CSU software,[49] which is available from the Protein Data Bank (PDB).[50] Native contacts between pairs of residues (i, j) with $|i–j| < 4$ were discarded from the native contact list because any three or four contiguous residues already interact through the angle and dihedral terms. The interface hydrophobicity was calculated based on the normalized occurrence of each amino acid in interfacial contacts, multiplied by its hydrophobicity factor.[51]

Binding Simulation Model

The Arc and lambda repressors and BS-RNase, (pdb codes 1arr, 1lmb, and 11ba, respectively) were also simulated with the Gō model,[41] which takes into account only interactions that exist in the native structure and

Table 1. Structural properties of the homodimers

Name	PDB code	interfacial NC[a] / monomeric NC	interfacial NC / no. of residues	monomeric NC / no. of residues	Interfacial hydrophobicity
Two-state dimers					
Troponin C site	1cta	1.19	1.50	1.26	0.53
Arc repressor	1arr	2.49	2.58	1.04	0.36
Gene V protein	2gvb	0.37	0.83	2.24	0.54
Factor for inversion stimulation	1f36	1.12	1.76	1.57	0.38
β nerve growth factor	1bet	0.45	1.06	2.26	0.41
GCN4 leucine zipper	2zta	1.79	1.61	0.90	0.42
C-JUN	1jun	1.70	1.47	0.86	0.46
TRP repressor[b]	2wrp	1.02	1.55	1.52	0.38
		1.30 ± 0.70	**1.54 ± 0.52**	**1.44 ± 0.55**	**0.43 ± 0.07**
Three-state dimers					
Lambda repressor	1lmb	0.27	0.60	2.21	0.39
Cro repressor	1cop	0.49	0.89	1.83	0.32
LFB1 transcription factor	1lfb	0.31	0.60	1.95	0.25
Streptomyces subtilisin inhibition	3ssi	0.29	0.73	2.56	0.33
Superoxide dismutase	1xso	0.13	0.40	3.02	0.41
Dihydroorotate Dehydrogenase	1dor	0.19	0.58	3.03	0.37
Methionine adenosyltransferase	1qm4	0.14	1.13	8.0	0.36
HIV-1 capsid	1A43	0.39	0.82	2.11	0.32
Triose phosphate isomerase	1tim	0.20	0.54	2.63	0.32
Sperm lysine	1lyn	0.22	0.49	2.24	0.29
Aspartate aminotransferase	1tar	0.17	1.22	7.08	0.29
Glutathione-S-transferase	1glq	0.29	0.40	1.36	0.26
Glutathione transferase	1gsd	0.34	0.87	2.58	0.25
		0.26 ± 0.10	**0.71 ± 0.25**	**3.12 ± 1.94**	**0.32 ± 0.05**

[a]NC = Native Contacts. [b]It is classified here as a two-state dimer, although a dimeric intermediate is detected during its binding, because its folding is coupled to the binding.

therefore does not include energetic frustration (or, alternatively said, only includes topological frustration). The Gō-model has already been used to study the folding of many monomers that fold in a 2-state fashion. In a survey of small proteins, Koga and Takada[52] showed that in more than half of the cases the structure of the transition state ensemble can be found using this model, which contains topological information alone. In some cases, symmetry between two possible transition state ensembles was broken by details of the pair interactions. In a more recent survey, Chavez et al.[53] found a correlation between the experimental folding rates and the rates (or the free energy barrier heights) obtained from the Gō simulations. In addition, the Gō model was successfully used to predict intermediates observed experimentally during the folding of several larger proteins.[54,55] However, while the structure and presence of partially folded intermediate ensembles is predicted well by these simple models and while they give a good correlation between the experimental and simulated folding rates, the absolute values of barriers and stabilities are sensitive to details of the potential. Indeed it seems that nonadditive potentials are needed to reproduce observed folding kinetics in full quantitative detail.[56]

We use here an off-lattice Gō model, where each residue is represented by a single bead centered on its α-carbon (C_α) position.[54] Adjacent beads are strung together into a polymer chain by means of a potential encoding bond length and angle constraints. The secondary structure is encoded in the dihedral angle poten-

tial and the nonbonded (native contact) potential. The interaction energy U at a given protein conformation Γ is given by

$$U(\Gamma,\Gamma_0) = \sum_{bonds}^{N-1} K_b (b_i - b_{0i})^2 + \sum_{angles}^{N-2} K_\theta (\theta_i - \theta_{0i})^2 + \sum_{dihedrals}^{N-3} \{K_\phi^{(1)}[1-\cos(\phi_i - \phi_{0i})] + K_\phi^{(3)}[1-\cos(3\times(\phi_i - \phi_{0i}))]\} + \sum_{\substack{native-contacts \\ |i-j|>3}} \left\{ \varepsilon \left[5\left(\frac{r_{0ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{ij}}\right)^{10} \right] \right\} + \sum_{\substack{non-native \\ contacts,|i-j|>3}} \varepsilon \left(\frac{C}{r_{ij}}\right)^{12} \quad (2)$$

In the equation, b_i, θ_i, and ϕ_i stand for the ith virtual bond length between ith and $(i + 1)$th residue, the virtual bond angle between $(i–1)$th and ith bonds, and the virtual dihedral angle around the ith bond, respectively. The parameters b_{0i}, θ_{0i}, and ϕ_{0i} stand for the corresponding variables in the native structure. In the framework of the model, all native contacts (as defined by the CSU) are represented by the 10–12 Lennard-Jones forms without any discrimination between the various chemical types of interaction. Moreover, both the intra- and intermonomeric contacts (interfacial contacts) are treated in the same way without any bias toward separate folding or toward binding. The r_{ij} and r_{0ij} are the C_α–C_α distances between the contacting residues i and j in conformation Γ and Γ_0 (the PDB structure), respectively. In the summation over non-native contacts, C (= 4.0 Å) parameterizes the excluded volume repulsion between residue pairs that do not belong to the given native contact set. In the paper, all temperatures and energies are reported in units of ε. For other parameters, we use similar values that have been used in several folding studies,[52–54,57] namely, K_b = 100.0, K_θ = 20.0, $K_\phi^{(1)}$ = 1.0, $K_\phi^{(3)}$ = 0.5, ε = 1.0.

To enhance the sampling of binding events, the two identical subunits of Arc and lambda repressors are linked by a polyglycine chain. This linker acts to hold the two unbound subunits (folded or unfolded) in close proximity during their motions; essentially the local concentrations are enhanced. The linker's length was determined by the distance between the C-terminus of subunit A and the N-terminus of subunit B. This length is sufficient to ensure the linker will not interfere with any intra- or intersubunit contacts that stabilize the folded dimer. To optimize its conformation with respect to the dimer, a minimization was performed on the linker including the two residues to which the linker is directly connected. Covalently linked Arc repressor[58] has been experimentally found to be fully functional with an enhanced folding rate and stability, suggesting indeed that the linker plays a passive, largely entropic role of keeping the unbound monomers at high local concentrations during folding. To further ensure the linker's role is only entropic, it has no nonbonded interaction (native contacts) with both subunits. All the parameters for the bonded terms of the linker residues were chosen to be smaller by one order of magnitude to enhance its flexibility and to reduce its energetic contributions. For BS-RNase a linker was not used in order not to introduce any effects on the swapping dynamics, which obviously requires a more subtle set of constraints due to potential tangling events. Instead, a constraint was applied preventing the center of mass distance of the two subunits from becoming greater than twice its value in the native complex.

In addition to the domain-swapped dimer of BS-RNase, another dimeric form in which there is no interchange has been experimentally observed.[44,59] Piccoli et al.[44] designated the domain-swapped dimer as M × M. The dimer formed by association of fully-folded monomers designated M = M. The interface of the M × M form of BS-RNase is five times larger than the interface in the M = M form. The interface of the M = M form constitutes the "secondary" interface of the domain-swapped form, M × M. With the aim of studying the conversion dynamics from monomer to dimer and also the thermodynamic properties of the two binding modes of BS-RNase, we conducted three types of simulations. The first type was designed to focus on the thermodynamics of the M = M form of BS-RNase, namely, only its native contacts are allowed in the Gō simulations. The same approach was also used focusing on the M × M form. In the third type of simulation, interactions present in both conformations are allowed and, accordingly, the Hamiltonian includes a Lennard-Jones potential for all the native contacts found in either the M = M or M × M forms.

Originally the structure of the M = M form was deduced through biochemical analysis[44] and has been recently observed by X-ray crystallography. Because a structure of M = M is not available, its structure was modeled based on the coordinates of the M × M conformation. This was done by interchanging the swapped helices in the M × M form, resulting in a structure with no swapping (i.e., the M = M form) and thus a much

smaller interface. In the M = M form the smaller interface is compensated by more compact monomers (i.e., more monomeric contacts). In the M = M form, the interfacial contacts, which exist also in the M × M form, are the outcome of docking between the two monomers. In principle, the two forms should have the same number of native contacts because the helices in their swapped or non-swapped conformations are in the same protein environment and form the same contacts. However, the M = M and M × M forms includes 642 and 651 native contacts, respectively. The additional contacts in the M × M form are due to different conformations of the hinges connecting the swapped regions to the rest of the protein.

For each of the three studied homodimers several constant-temperature molecular dynamics simulations were performed (using the simulation package AMBER6 as an integrator[60]) starting from either the dimeric conformation or the unfolded monomers. The set of trajectories was combined using the Weighted Histogram Analysis Method (WHAM)[61] to provide the transition temperatures from the peaks of the specific heat versus temperature and to calculate thermodynamic properties of the systems. The free energy surface of a binding process is projected onto several candidate reaction coordinates for folding and binding: the fractions of monomeric native contacts, interfacial native contacts, the total number of native contacts, and the distance between the center of mass of the two subunits. In the free energy calculations, the energy terms associated with the linker residues were not taken into account to enable a comparison between a dimer and an isolated monomer folding.

RESULTS

Thermodynamic Analysis of Homodimer Subunits

One of the main goals of our current study was to survey a set of dimeric proteins with analytical landscape theory, focusing on those whose association mechanism has been experimentally studied, using structural and thermodynamic analysis. If the monomeric chains that constitute the dimer complex fold quickly and reliably in the absence of their respective partner chain, then it is most likely that binding events occur after folding, i.e., there is no coupling between binding and folding. On the other hand, if an isolated monomeric chain is largely unstructured prior to binding, then folding would occur concomitantly with binding, caused either by energetic stabilization due to favorable interface contacts or some kinetic speedup due to coupling of binding and folding. Given these scenarios, it becomes important to determine whether a monomeric chain is thermodynamically stable in the absence of a partner chain. As described in the Models and Methods section, we have used the Minimally Frustrated Random Energy Model (MFREM) to compute the free energies of folding for all chains that are part of 2-state and 3-state protein complexes considered in this paper. We should emphasize here that MFREM theory is usually applied to describe gross features of energy landscapes of an ensemble of proteins, while in this work we use it to estimate the stabilities of individual proteins. In addition, with the coarse-grained inter-residue potential used to compute parameters for the MFREM, we expect to find only qualitative trends in stability when averaging over groups of proteins.[22, 40]

In Fig. 1, the free energies of folding for monomers forming 2-state dimers (red bars), 3-state dimers (blue bars), and 210 reference monomer chains (gray spheres) are shown as a function of chain internal hydrophobicity and the average density of internal contacts per chain residue. First, we observe that the horizontal coordinates clearly separate monomeric chains belonging to the 2-state and 3-state dimers into two distinct groups. Monomers in which dimers have a 3-state folding mechanism are internally more hydrophobic than monomers where dimers are formed via a 2-state folding mechanism. The monomers of 3-state dimers have a higher density of contacts per residue than those of monomers of 2-state dimers. Notice that 210 reference monomers that are known to fold independently are found clustered in the region where 3-state dimers are located. Although these observations are in accord with naive intuition, it is still remarkable that just using structural characterization (i.e., without sequence data) of naturally occurring monomers alone makes it possible to predict whether the folding process for that particular monomer requires coupling to a binding process. Clearly, the minimal frustration constraint is needed for this result. A single amino acid mutation can surely destabilize a monomer, so if the evolutionary bias is not included, sequence information would be needed to predict the mechanism.

Amazingly, the crude contact energy model can predict stabilities of monomeric proteins. Monomers of dimers with a 3-state folding mechanism are more stable individually than monomers that fold upon binding. Thus, as far as group trends go, the MFREM theory coupled with the coarse-grained inter-residue pair potential make a qualitatively correct postdiction. In particular, five out of seven 2-state monomeric chains are predicted to be unstable, while only one (2gvb) is computed to be slightly stable. One 2-state protein (1bet), however, is predicted to be extremely stable. It turns out to be in fact the most stable monomeric chain among all

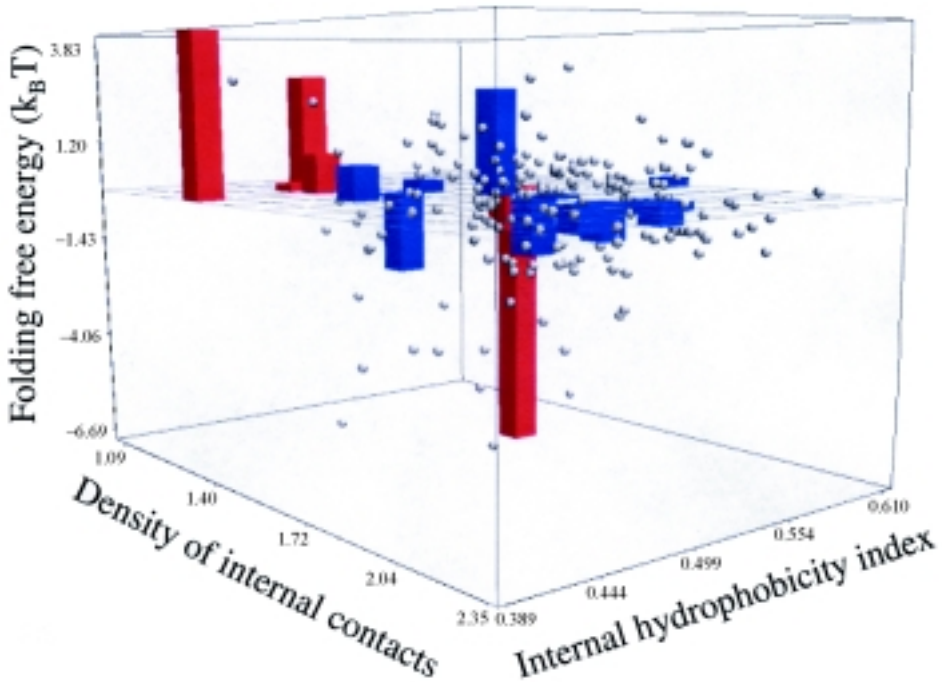

Fig. 1. Free energies of folding computed with the MFREM model, as described in Models and Methods section. The 2-state dimers are indicated with red bars, the 3-state dimers, with blue bars. The 210 reference monomers are indicated with gray spheres. The hydrophobicity index is computed based on residues that form contacts, by unweighted averaging over residues using the Pacios hydrophobicity scale.[51] Density of contacts is defined as number of internal contacts per monomer.

monomers considered, including the 210 reference proteins. Although we may not exclude the possibility that this is a spurious result, we suggest on the contrary that 1bet (β nerve growth factor) is indeed thermodynamically stable. Nevertheless, it has been observed to fold concomitantly with dimerization,[62] in disagreement with this suggestion. We must entertain the possibility that folding of monomeric 1bet is arrested due to a high folding barrier. This will not occur in the pairwise additive Gō model, but may arise in the minimally frustrated nonadditive model owing to the large solvent-accessible surface of the protein.[26] Thus binding perhaps catalyzes folding. Other experiments have observed a compact 1bet monomeric intermediate at acidic pH.[62] This would agree with our calculations. In the simulations of 1bet dimer binding and folding using the Gō Hamiltonian, a kinetic intermediate was actually detected with partially native contacts.[31]

The composition of binding interfaces is also expected for folded chains that bind from those where coupling of binding to folding occurs. The number of monomeric and interfacial native contacts as well as the monomeric and interfacial hydrophobicity were calculated for each dimer. Figure 2A shows a "phase diagram" correlating the experimental association mechanism with a structural classification of 2- and 3-state homodimers based both on the number of intramonomeric and interfacial native contacts and on the hydrophobicity of the interface.[31] 2-state dimers are characterized by a higher ratio of interfacial contacts to monomeric contacts. This arises from both an extensive interface and lower number of monomeric contacts (Table 1). 2-state dimers are less compact and have larger interfaces than 3-state dimers.[26,31] 2-state homodimers also have a more hydrophobic interface. This agrees with previous structural analyses,[63,64] that suggest that when folding and association are coupled the resulting interface is hydrophobic, similar to the core of a single domain protein. For hydrophilic interfaces, association follows folding and is perhaps steered and guided by long-range electrostatic interactions.[65,66] Some of our recent work suggests that residue–residue interactions at a longer separation, mediated by water, play an important role in binding recognition for complexes having hydrophilic interfaces.[40]

The free energy for folding of the individual monomers that constitute each homodimer is shown in Fig. 2B with the same axes used for the "phase diagram" in Figure 2A. One may expect that the subunits of the 2-state homodimers will have positive values of free energy for folding, as they are natively unstructured on

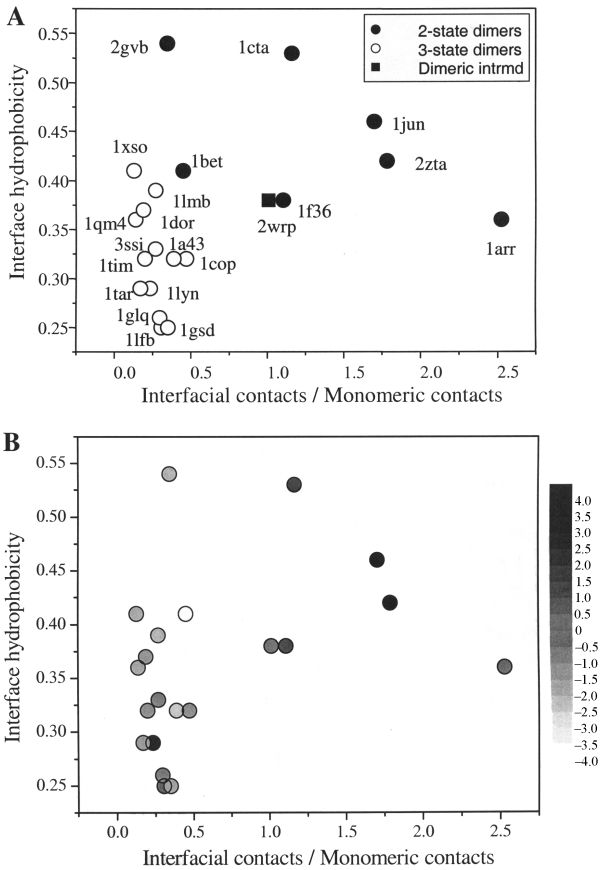

Fig. 2. A "phase diagram" that correlates the association mechanism of the homodimers with their structural properties. The 2- and 3-state homodimers are structurally classified based on the number of intramonomeric and interfacial native contacts as well as the interface hydrophobicity. The classification as 2-state or 3-state is based on experimental data (A) and on the free energy stability for monomer folding (B). In general, a 2-state dimer is characterized by (i) a higher ratio of interfacial native contacts to monomeric native contacts, (ii) a more hydrophobic interface in comparison to a 3-state dimer (see Table 1), and (iii) a less stable monomer. 2wrp (Trp repressor) is a 3-state homodimer with a dimeric intermediate. It is denoted using the same color as the 2-state dimers since its dimerization does not involve a pre-existing folded monomer.

their own, and that the monomers that compose the 3-state homodimers will have negative values. It is shown that higher values of free energy for monomer folding describe the 2-state rather than the 3-state dimers. Although not all the studied 2(3)-state dimers are characterized by positive (negative) values of free energy for folding, presumably, due to the theory simplicity, yet a clear distinction is observed between two types of dimers.

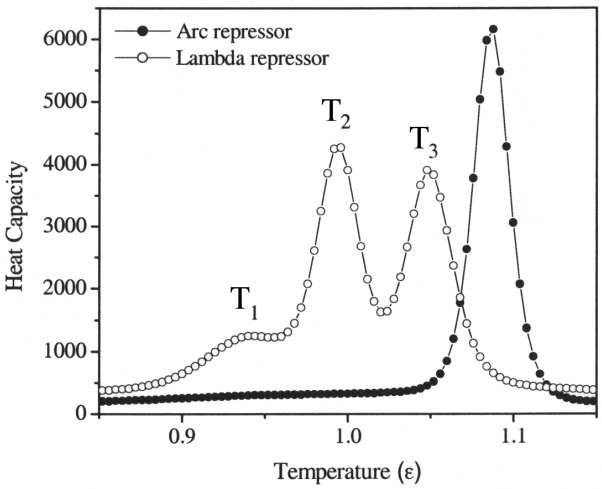

Fig. 3. The heat capacity of Arc and lambda repressors as a function of temperature. The heat capacity profile of Arc repressor includes a single peak and that of lambda repressor includes three peaks (designated $T_1 - T_3$).

Binding Mechanisms: Folding by Binding or Vice Versa

We simulated the formation of dimeric Arc repressor (2-state dimer) and lambda repressor (3-state dimer) using a simple Gō model. This model captures solely topological frustration. During constant-temperature simulations, we monitored the number of native contacts, Q_{Total}, which is a sum of monomeric native contacts (Q_A and Q_B) and interfacial native contacts ($Q_{Interface}$). Folding and binding rate coefficients are strongly temperature dependent. It is thus crucial to accurately determine the transition temperature. For this purpose we apply the WHAM analysis[61] to compute the heat capacity($C_v = (\langle E^2\rangle - \langle E\rangle^2)/kT^2$) as a function of temperature. The heat capacity plots for the association reactions of the two dimers are shown in Fig. 3. The single peak observed in the specific heat curve of Arc repressor indicates, in agreement with experiment, a cooperative transition from the unfolded monomeric chains to a folded dimer. On the other hand, the three peaks in the specific heat curve of lambda repressor point to a more complicated binding mechanism that includes three transition temperatures ($T_1 - T_3$). The free energy profiles for dimerization are shown in Fig. 4 and indicate different binding mechanisms. The free energy profile for Arc repressor, in harmony with experiment, includes two states: a minimum consisting of two unfolded chains and a minimum for the folded dimer. The barrier between them is about 6ε (Fig. 4A). The free energy profile of lambda repressor exhibits four minima. In addition to the states that correspond to unfolded

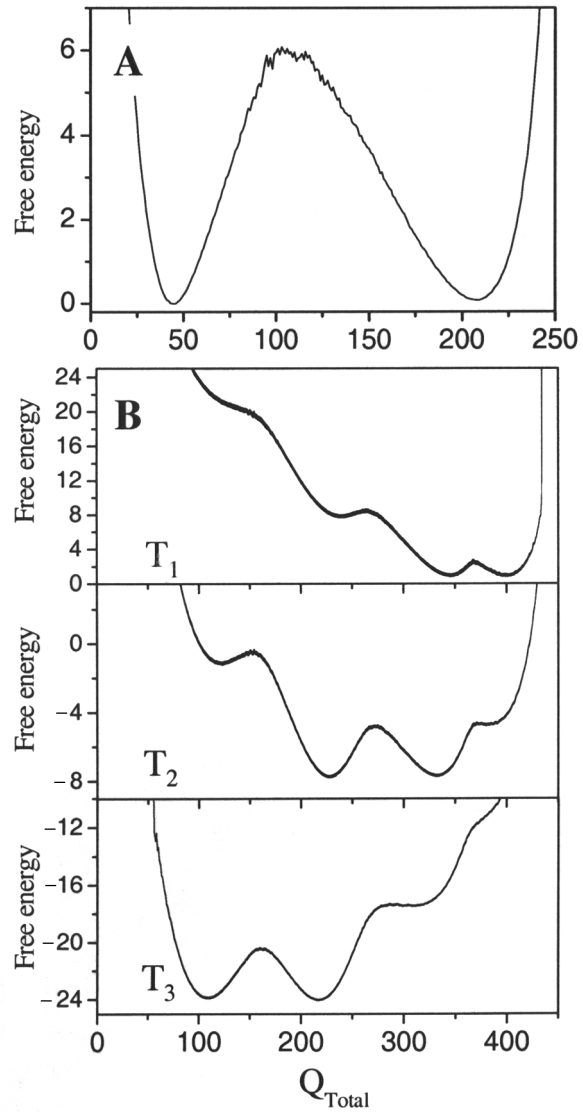

Fig. 4. Free energy as a function of the reaction coordinate Q_{Total} for Arc and lambda repressors. The free energy curve of Arc repressor (A) was calculated at its folding temperature (1.08ε) (see Fig. 3). For lambda repressor (B) the free energy curves are plotted at all three transition temperatures, T_1, T_2, and T_3, (0.94ε, 0.99ε, and 1.05ε, respectively) obtained as peaks in the heat capacity curve (Fig. 3).

chains and a folded dimer, there are other states in which only a single chain is folded (while the other is unfolded) and one in which there are two folded chains that are not bound to each other (Fig. 4B). The free energy profiles are plotted at the three transition temperatures reflecting the three equilibria between the states. For example, T_1 is the transition temperature for binding, and it is seen that at this temperature two folded noninteracting subunits (i.e., high Q_A and Q_B, but low $Q_{Interface}$) have the same free energy as a folded dimer (at temperatures lower than T_1 the dimer becomes the most stable state).

The free energy surfaces of Arc and lambda repressors projected along various reaction coordinates (Q_A, Q_B, $Q_{Interface}$, Q_{Total}, and the distance between the center of mass of the two subunits) are shown in Fig. 5. The free energy surfaces projected along Q_{Total} and the separation distance between the two subunits (Fig. 5A,B) suggest different association mechanisms for the two dimers. Starting from two unfolded chains, which can be far away from each other (the distance between their center of mass is up to 100 Å), folding and binding proceed directly to a folded dimeric Arc repressor, where the separation distance between the two monomers becomes very restricted. In the case of lambda repressor there are intermediates. The surfaces projected along the Q coordinates (Fig. 5C–F) demonstrate the existence of coupling between the folding of the two subunits that constitute both Arc and lambda repressors as well as the coupling between folding and binding. The folding of two chains of Arc repressor are coupled, and no folded monomer is observed independently of the other subunit. In lambda repressor the two chains are autonomous entities and can fold regardless of the presence of the other subunit and do not require the interface formation.

Analysis of the transition state (TS) ensemble also reveals the different association mechanisms of the two homodimers. Figure 6 shows the fraction of time that individual native contacts are formed at the TS ensemble. The TS ensemble for binding (and folding) of Arc repressor is located in the region $90 < Q_{Total} < 130$, while the ensemble for lambda repressor has $360 < Q_{Total} < 380$. Comparing the probability of native contact formation at the TS ensemble and at the native structure indicates that partial secondary structure elements exist in the monomeric Arc repressor while the monomers of lambda repressor are fully folded and the only partially formed contacts at the binding TS of lambda repressor are interfacial contacts (indicated by dashed ellipse). In the TS ensemble of Arc repressor, the intramolecular interactions are on the average more formed relative to intermolecular interactions. This agrees with experimental characterization of the TS ensemble of Arc repressor using mutational analysis (known as Φ-value analysis), which indicates that the monomers fold slightly before binding.[67,68]

Lambda repressor forms by association of already-folded monomers. To elucidate whether its binding obeys a lock-and-key, induced-fit, or conformational selection mechanism, a trajectory simulated at T_1, which includes eight binding/unbinding events, was analyzed (Fig. 7). All the conformations of the two subunits (bound and

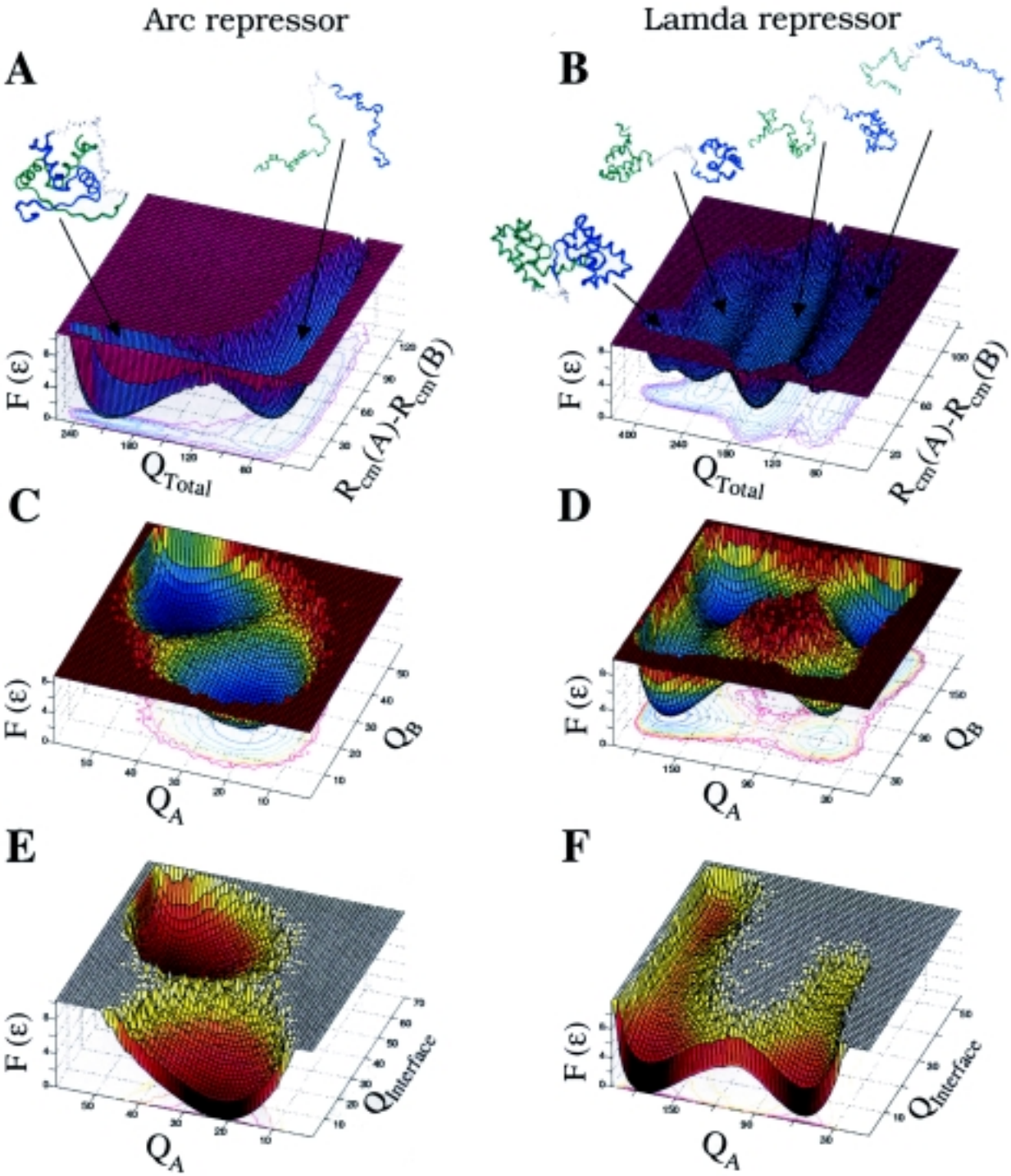

Fig. 5. Free energy surfaces for folding and binding of Arc repressor (2-state dimer) and of lambda repressor (3-state dimer). Free energy surfaces of the simulated homodimers are plotted as a function of the intrasubunit native contacts (Q_A and Q_B), intersubunit native contacts, ($Q_{Interface}$), the total number of native contacts (Q_{Total}), and the separation distance between the two chains (Rcm(A)–Rcm(B)). The simulations reproduce the experimentally inferred mechanisms regarding the coupling between folding and binding. The free energy surfaces of Arc and lambda repressors are calculated at 1.08ε and 0.99ε, respectively.

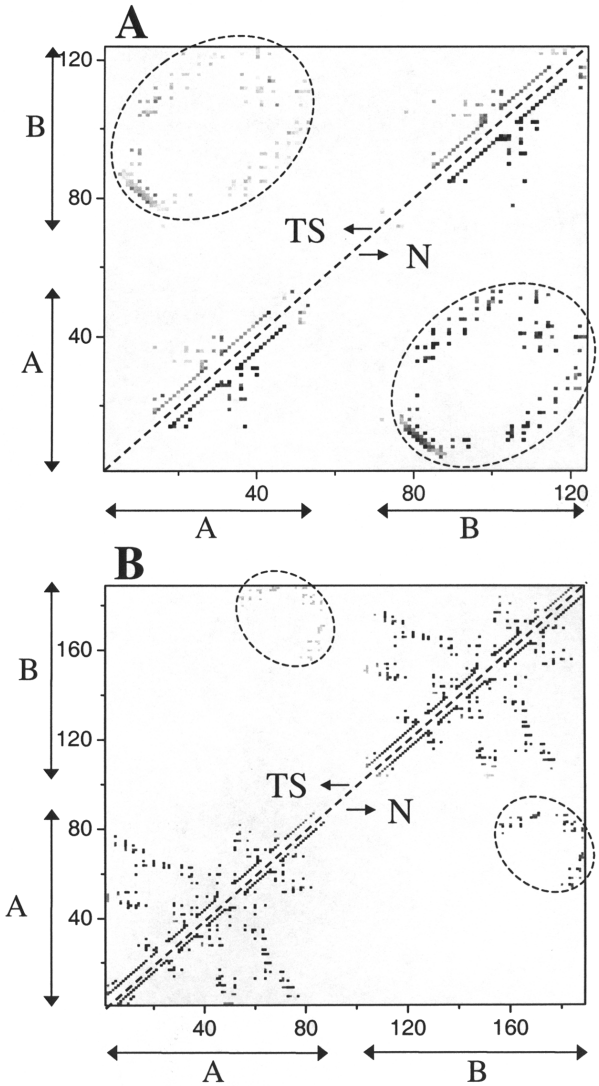

Fig. 6. The probability of native contact formation at the binding transition state as well as at the native states of Arc repressor (A) and lambda repressor (B) is shown as a contact map. The probabilities are shown with a gray scale (zero probability is indicated by white and fully formed contact by black) for native contacts in monomers A and B and the interfacial contacts (indicated by an ellipse). Residues 54–71 of Arc repressor and 88–102 of lambda repressor, respectively, are linkers connecting the C terminus of subunit A to the N terminus of subunit B. These residues have no nonbonded interactions in the simulation.

unbound) were analyzed using principal component analysis.[69,70] The separation between the conformations that participate in the encounter complex of different binding events indicates that the binding follows an induced-fit mechanism. Each binding event is initiated by a different conformation and not by a pre-selected frozen conformation with the optimal interface for recognition, as suggested by the lock-and-key and the conformational selection mechanisms. Thus, it is evident that the stable complex is obtained by optimizing the interface formed initially in different encounter complexes.

Association by Domain-Swapping Mechanism

Many proteins assemble via the domain-swapping mechanism.[12,71] In these oligomers, in accord with the minimal frustration principle, the swapped domain makes identical non-covalent bonds with neighboring domains in the closed monomer and in the oligomer. These interactions constitute the "primary" interface.[71,72] As the subunits are often close to each other in a domain-swapped oligomer, a new interaction interface that is absent in the monomer may be formed, and this is termed the "secondary" interface. As oligomers are entropically disfavored, the stability of the "secondary" interface (and the hinge loop that connects the exchanging subunit with the rest of the protein) will determine whether domain-swapping will be thermodynamically favorable.[73,74] Whether it is the monomeric or the domain-swapped oligomer form that is more stable, the two forms are expected to be separated by a high-energy barrier. This barrier arises from a requirement for partial unfolding, which will result in disrupting non-covalent bonds to enable interconversion. In some cases (e.g., p13suc1[72,73]) a complete unfolding of the monomers is an obligatory step for a domain-swapped dimer to be formed.

We illustrate these notions with the dimerization of monomeric BS-RNase. The BS-RNase is a good candidate for studying the mechanism of converting a monomeric protein into a domain-swapped oligomer because both of its quaternary structures are structurally characterized:[44] a domain-swapped dimer, M × M, and a dimer with no interchange, M = M (see Fig. 8). It has been suggested that the simultaneous existence of two quaternary forms of the same biomolecule is a sign of evolution in progress.[59,75,76] The swapped dimer, M × M, has been found to be more functional, while the M = M dimer apparently is not functional. It was therefore speculated that the M × M form will eventually be selected by evolution as the only stable form. To study binding by domain-swapping and to examine the evolutionary question, three types of Gō simulations were performed. In two simulations only the native contacts that stabilize either the M = M or M × M forms were introduced. In another there are contacts that occur in either form with the same weight (see Models and Methods). The first two types of simulations provide a good route to the thermodynamic properties of binding via the domain-swapping mechanism relative to the

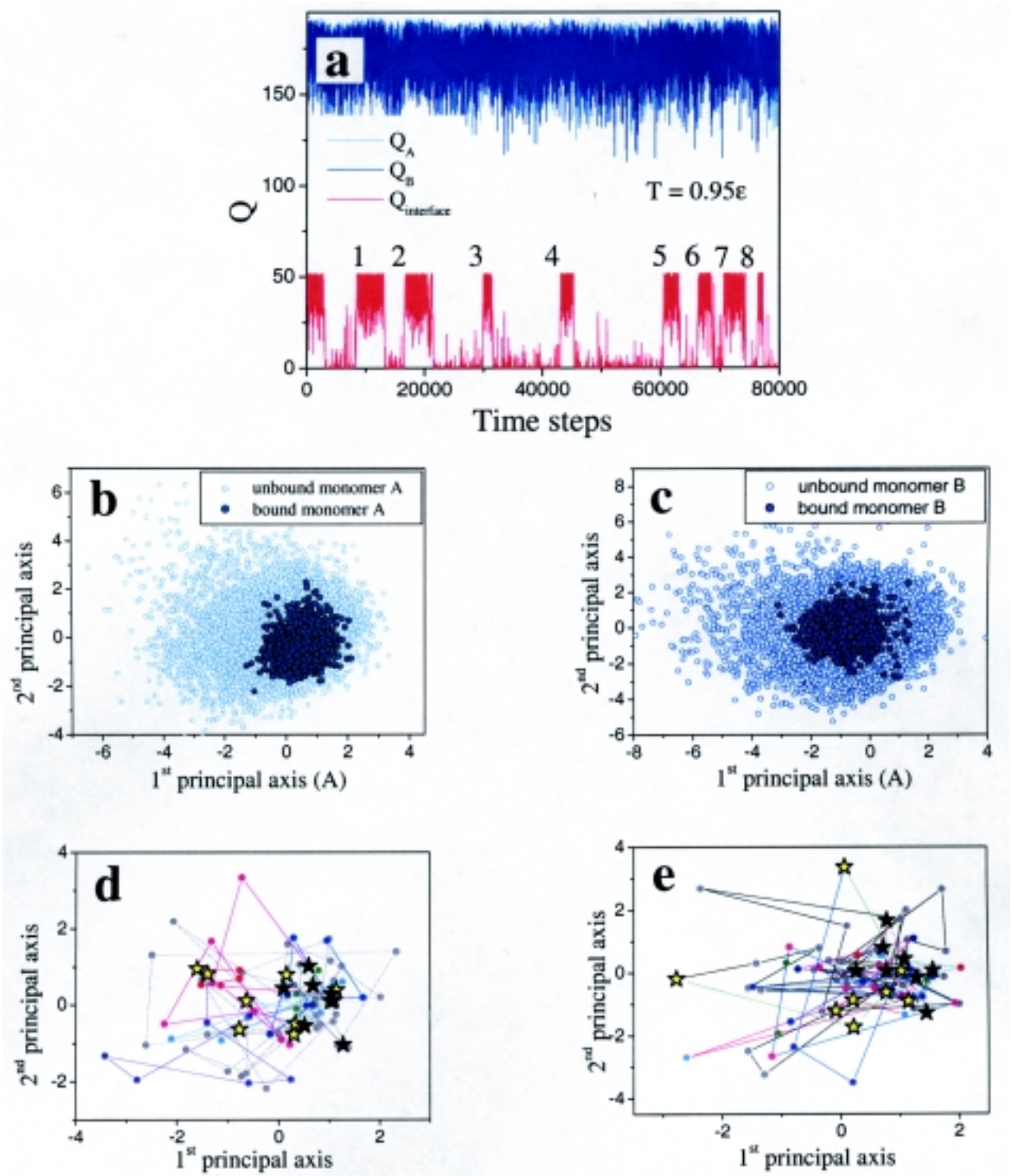

Fig. 7. An illustration of "induced fit" binding during the association of lambda repressor. (**a**) A trajectory of lambda repressor at T_1 (= 0.94ε) shows eight association events between two already folded monomers. All the conformations of subunit A (**b**) and B (**c**) are projected separately on the two principal axes. The empty and full circles are for unbound and bound subunits, respectively, indicating a reduced flexibility due to binding. The bottom panels illustrate the conformations of monomers A and B (**d** and **e**, respectively) along each event starting with a conformation with no interfacial native contacts (yellow star) and ending with a conformation with maximal interfacial native contacts (black star). This illustrates that each binding event is initiated by different conformations and not by a single pre-selected conformation with the optimal interface for recognition. Accordingly, it is evident that the stable complex can be obtained from different encounter complexes with a weak interface.

assembly of folded monomers. The third type of simulation represents a more realistic scenario where a competition between the two forms is possible and can be kinetically accessed.

The heat capacity profiles for the binding of two monomeric BS-RNase molecules to form either M = M or M × M forms are shown in Fig. 8. For the two binding modes (association between already folded monomers and domain-swapping) three peaks were observed, indicating, as in the case of lambda repressor (Fig. 3), the existence of three transitions. Although the two conformations of BS-RNase share almost the same number of native contacts, the position of the three peaks in their heat capacity profiles are different. These differences originate from the dimers having both different monomeric structures and different size interfaces between the two monomers. The transition temperatures that correspond to monomer folding (T_2 and T_3) are similar, with a larger specific heat for the non-swapped dimer due to its larger number of monomeric contacts. Most significantly, the binding temperature (T_1) of the two different dimers of BS-RNase is far from the folding temperatures associated with the folding of a single chain (T_3) and of the two chains (T_2). This is due to different interface sizes of the two dimeric forms that dictate different binding temperatures. For the M = M form of BS-RNase there is a clear separation between folding and binding, while for the domain-swapped dimer, M × M, folding and binding occur at closer temperatures, suggesting that coupling between folding and binding is more probable in this case.

The free energy surfaces projected onto Q_A, Q_B, and $Q_{Interface}$ for the two dimeric conformations of BS-RNase are shown in Fig. 9A–D. The domain-swapped dimer contains a larger interface ($Q_{Interface} < 110$) but it also has fewer monomeric contacts (Q_A, $Q_B < 280$) than in the M = M form. The α-helices exchange to form the M × M form of BS-RNase occurs after the two subunits are folded. This contrasts with the monomer unfolding, which is required for the formation of domain-swapped p13suc1.[72,73] The different swapping mechanisms originate from unequal participation of the exchanged region in the protein core (14% and 31% for BS-RNase and p13suc1, respectively). During the association of two monomers via the domain-swapping mechanism, an intermediate with a partial interface where only a single α-helix is exchanged was detected in the simulation (Fig. 9D). Having found the thermodynamic and the dynamic properties of the binding as either the M = M or M × M form, we next studied the swapping mechanism allowing competition between the two dimeric forms in the same model. When contacts of both structures are allowed in the Gō model two native monomeric structures are possible, which can in principle bind and result in either M = M or M × M conformation. As we are interested in understanding the mechanism of domain-swapping, the simulations were performed at the temperature T_1, where swapping was observed for the

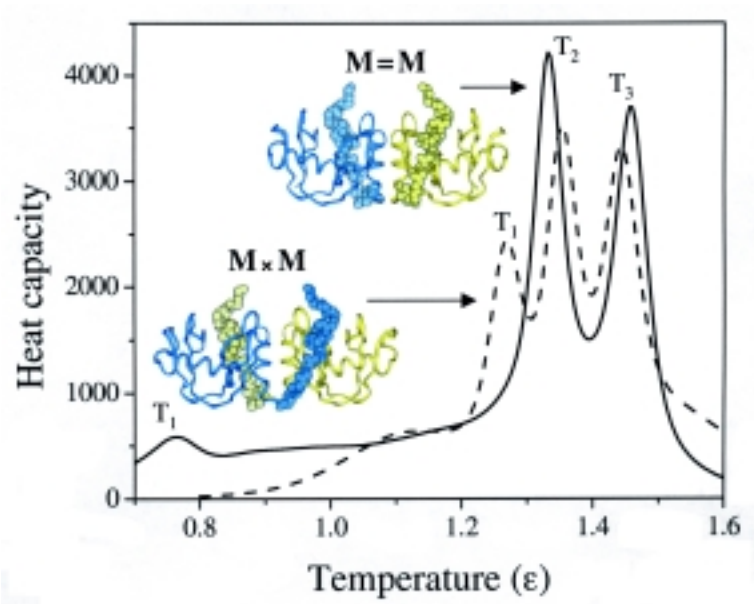

Fig. 8. The heat capacity for the two quaternary forms of dimeric BS-RNase: an exchanged dimer named M × M and a structure with no exchange named M = M. Three peaks are found in the heat capacity curves of both forms, indicating similar folding and binding transitions. While similar transition temperatures are found for the monomer folding (T_2 and T_3) of the two forms, distinct binding transition temperatures (T_1) of the two forms reflect the different size and geometry of their interfaces.

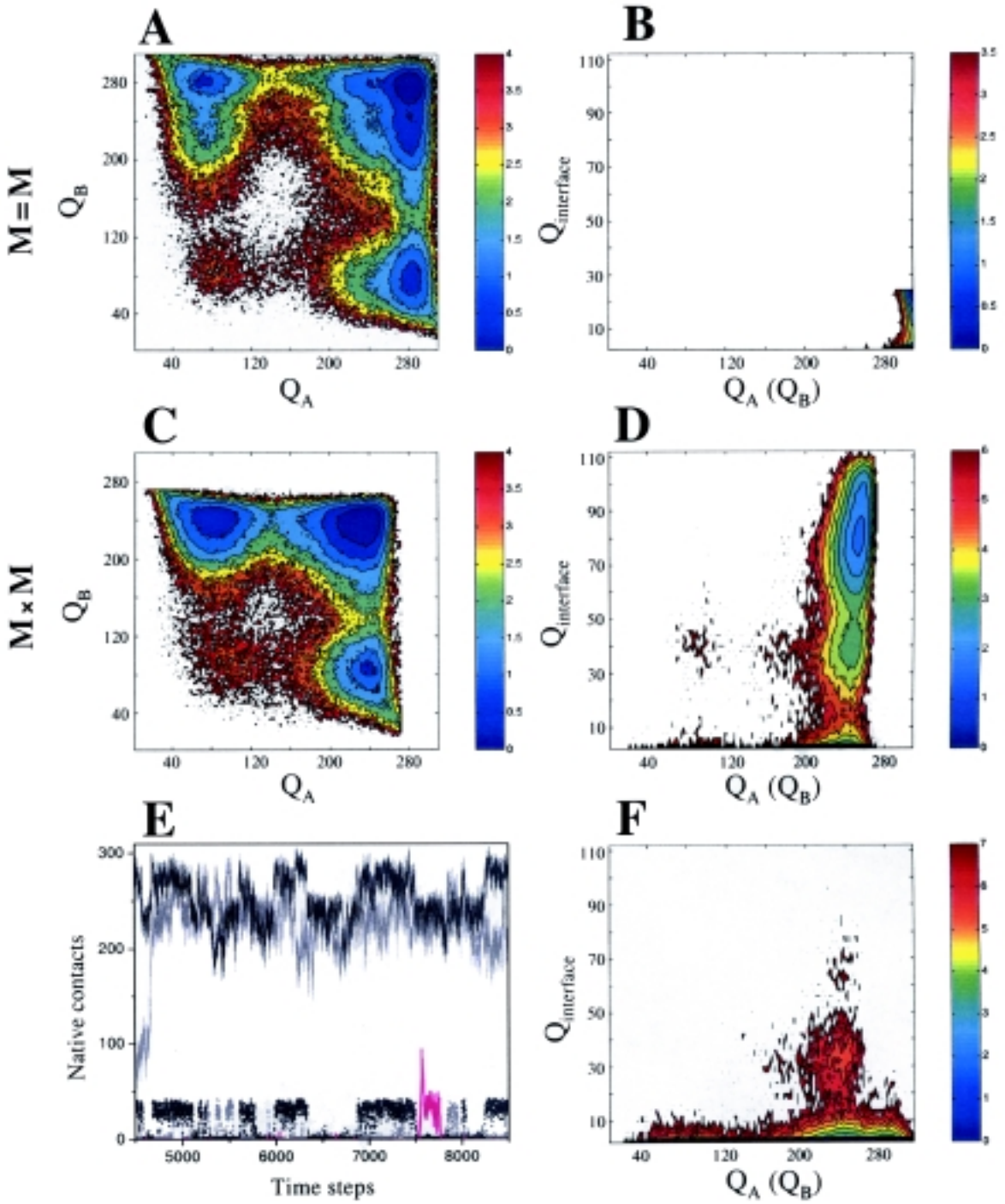

Fig. 9. Free energy surfaces for folding and binding BS-RNase in its two quaternary forms: M = M (A and B) and M × M (C and D). The free energy surfaces for each form of BS-RNase were obtained by Gō simulation where only the native contacts included in the form under consideration were allowed. Accordingly in the simulations of the M = M form, 642 contacts were taken into account (310 × 2 monomeric contacts + 22 interfacial contacts), and in the simulations of the M × M form, 651 contacts were allowed (270 × 2 monomeric contacts + 111 interfacial contacts). A trajectory of the simulation that includes all the contacts in both M = M and M × M forms (E) illustrates a domain-swapping event. The monomeric contacts in monomers A and B are shown in gray (light and dark, respectively), the contacts of the swapped helix in the non-swapped structures are shown by dots (light and dark gray), and the interfacial contacts in pink. A projection of the 64 trajectories where the contacts in both M = M and M × M forms were allowed (F) illustrates a combination of (B) and (D) with much less sampling of swapping events (high $Q_{Interface}$).

model that includes M × M contacts alone (see Fig. 8). Sixty-four trajectories were simulated (each includes over 4×10^7 integration time steps), and only two binding events to form M × M were observed (the M = M form was not detected, although it is allowed since it is stable at much lower temperatures). A trajectory where binding occurred via the domain-swapping mechanism is shown in Fig. 9E. Although we cannot yet estimate the barrier that separates the two dimeric forms, the fact that the sampling of the domain-swapped dimer was significantly affected by the competition between the monomeric and domain-swapped dimer suggests that interchange is a rare event.

CONCLUSIONS

The dimerization of protein chains was studied here both analytically and by simulations using a simplified model. Two sets of homodimers were selected based on their experimental classification regarding the existence of monomeric intermediates during the binding process. One set includes 2-state homodimers whose monomers are natively unstructured. The folding of these dimers is coupled directly with binding. The 3-state homodimers, which constitute the other set, have monomers that are stable on their own and their folding occurs before binding. Using the Minimally Frustrated Generalized Random Energy Model, the free energies for folding of the monomers of the 2- and 3-state homodimers were calculated and compared to that of a large set of known folded single-domain proteins. On average, higher values of the free energy for folding were found for the monomers of 2-state dimers, supporting their experimental classification as intrinsically unstructured proteins. These monomers were found to be less compact and less hydrophobic than folded proteins (i.e., the monomers of the 3-state homodimers and stable single-domain proteins) indicating that the existence of coupling between folding and binding can be obtained based on thermodynamic analysis of the complex monomers. In addition, the interfaces of 2-state homodimers are more hydrophobic than those formed by assembly of already folded subunits, reflecting again that the binding between two unfolded proteins has many similarities to protein folding.

The dynamics of protein association both where monomer folding occurs prior to or is coupled to recognition was studied by simulating the formation of the dimeric Arc repressor (2-state dimer) and lambda repressor (3-state dimer) using a simple model that takes into account only native contacts. This model does not include energetic frustration and thus corresponds to a perfectly funneled energy landscape. Our simulations reproduced the experimental observations regarding the role of folded monomers in the association. The agreement between the binding mechanisms found in experiment and from simulations with energetically minimally frustrated models strongly argue for the notion that binding processes have funneled landscapes.[31] Binding on funneled surfaces is robust to mutation and is mainly governed by the structure of the network of contacts in the monomer and at the interface (i.e., the protein topology).

The domain-swapping mechanism for protein assembly is an efficient binding mechanism that overcomes the need for separate interface design by taking advantage of the minimal frustration principle using interfacial contacts to bind the molecules that are identical to those found to stabilize the monomers before the exchange takes place. The domain-swapping mechanism was studied for bovine seminal ribonuclease, which forms two different quaternary structures: one swapped dimer and the other a dimer with no swapping, which thus has a much smaller interface. The thermodynamic properties of each form were obtained by simulations in which a single form was allowed. It was shown that a coupling between folding and binding is more likely for the domain-swapped dimer of bovine seminal ribonuclease than for the dimer with no swapping. The kinetic accessibility of domain-swapped dimer significantly decreases when the competition between the two forms is introduced, reflecting that the interchanging requires crossing a high barrier. Understanding how binding occurs when secondary structure elements interchange is essential, as there are many examples of this interchange in protein assembly. Interchange may also be important in pathological aggregation. In principle, any protein can interact with another copy of itself by exchanging identical domains with partners while still respecting the minimal frustration principle. The fact that pathologic aggregates usually involve a single protein provides strong support for this principle. Likewise the observation of dimeric intermediates in many folding experiments at high concentration, as emphasized by Oliveberg,[77] is evidence for funneling. Finding a quantitative criterion for the avoidance of such domain-swapped dimers for proteins that should not form aggregates will be extremely valuable.

Here we have shown that the ability of globular proteins to assemble themselves into clusters with well-defined structures can be understood through energy landscape theory. The comparison between experiments and simplified simulations should be extended in the future to include detailed microscopic analyses, especially on the effect of mutations on the binding mechanism. The Gō model can be extended to include non-native interactions to quantify the degree of energetic

frustration in binding and the extent to which proteins are designed for efficient binding. In addition, including non-native interactions in such simulations will allow an examination of the degree of specificity required in biomolecular assembly within living cells.

Acknowledgments. 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. Y.L. also acknowledges the support from the Rothschild and Fulbright foundations. G.P. is supported through a National Institutes of Health postdoctoral fellowship award. Computations were carried out at the UCSD KeckII computing facility (partially supported by NSF-MCB).

REFERENCES AND NOTES

(1) Gavin, A.-C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M. *Nature* **2002**, *415*, 141–147.
(2) Ho, Y.; Gruhler, A.; Heilut, A.; Bader, G.D.; Moore, L.; Adams, S.-L. *Nature* **2002**, *415*, 180–183.
(3) Salwinski, L.; Eisenberg, D. *Curr. Opin. Struct. Biol.* **2003**, *13*, 377–382.
(4) Janin, J.; Seraphin, B. *Curr. Opin. Struct. Biol.* **2003**, *13*, 383–388.
(5) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. *Nature* **2001**, *411*, 41–42.
(6) Dobson, C.M. *Trends Biochem. Sci.* **1999**, *24*, 329–332.
(7) Fischer, E. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985–2991.
(8) Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R. *Protein Eng.* **1999**, *12*, 713–720.
(9) Bosshard, H.R. *News Physiol. Sci.* **2001**, *16*, 171–173.
(10) Koshland, D.E.J. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–123.
(11) Bennett, M.J.; Schlunegger, M.P.; Eisenberg, D. *Protein Sci.* **1995**, *4*, 2455–2468.
(12) Liu, Y.; Eisenberg, D. *Protein Sci.* **2002**, *11*, 1285–1299.
(13) Bryngelson, J.D.;Wolynes, P.G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524–7528.
(14) Leopold, P.E.; Montal, M.; Onuchic, J.N. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8721–8725.
(15) Onuchic, J.N.; Wolynes, P.G.; Luthey-Schulten, Z.; Socci, N.D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3626–3630.
(16) Taverna, D.M.; Goldstein, R.A. *J. Mol. Biol.* **2002**, *315*, 479–484.
(17) Neet, K.E.; Timm, D.E. *Protein Sci.* **1994**, *3*, 2167–2174.
(18) Xu, D.; Tsai, C.-J.; Nussinov, R. *Protein Sci.* **1998**, *7*, 533–544.
(19) Shakhnovich, E.I. *Nat. Struct. Biol.* **1999**, *6*, 99–102.
(20) Wright, P.E.; Dyson, H.J. *J. Mol. Biol.* **1999**, *293*, 321–331.
(21) Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; Ausio, J.; Nissen, M.S.; Reeves, R.; Kang, C.; Kissinger, C.R.; Bailey, R.W.; Griswold, M.D.; Chiu, W.; Garner, E.C.; Obradovic, Z. *J. Mol. Graph. Modeling* **2001**, *19*, 26–59.
(22) Papoian, G.A.; Wolynes, P.G. *Biopolymers* **2003**, *68*, 333–349.
(23) Dyson, H.J.; Wright, P.E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
(24) Kirwacki, R.; Hengst, L.; Tennant, L.; Reed, S.; Wright, P. *Proc. Natl. Acad. U.S.A.* **1996**, *93*, 11504–11509.
(25) Spolar, R.; Record, M. *Science* **1994**, *263*, 777–784.
(26) Gunasekaran, K.; Tsai, C.-J.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28*, 81–85.
(27) Shoemaker, B.A.; Portman, J.J.; Wolynes, P.G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868–8873.
(28) Janin, J.; Henrick, K.; Moult, J.; Eyck, L.T.; Sternberg, M.J.E.; Vajda, S.; Vakser, I.; Wodak, S.J. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 2–9.
(29) Ben-Zeev, E.; Berchanski, A.; Heifetz, A.; Shapira, B.; Eisenstein, M. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 41–46.
(30) Mendez, R.; Laplae, R.; Maria, L.D.; Wodak, S.J. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 51–67.
(31) Levy, Y.; Wolynes, P.G.; Onuchic, J.N. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 511–516.
(32) Onuchic, J.N.; Luthey-Schulten, Z.; Wolynes, P.G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 539–594.
(33) Onuchic, J.N.; Socci, N.D.; Luthey-Schulten, Z.; Wolynes, P.G. *Folding and Design* **1996**, *1*, 441–450.
(34) Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181–1190.
(35) Zhang, C.; Chen, J.; DeLisi, C. *Proteins: Struct., Funct., Genet.***1999**, *34*, 255–267.
(36) Tovchigrechko, A.; Vakser, I.A. *Protein Sci.* **2001**, *10*, 1572–1583.
(37) Kumar, S.; Ma, B.; Tsai, C.-J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10–19.
(38) Verkhivker, G.M.; Bouzida, D.; Gehlhaar, D.K.; Rejto, P.A.; Freer, S.T.; Rose, P.W. *Curr. Opin. Struct. Biol.* **2002**, *12*, 197–203.
(39) Wang, J.; Verkhivker, G.M. *Phys. Rev. Lett.* **2003**, *90*, 188101.
(40) Papoian, G.A.; Ulander, J.; Wolynes, P.G. *J. Am. Chem. Soc.* **2003**, *125*, 9170–9178.
(41) Ueda, Y.; Taketomi, H.; Go, N. *Biopolymers* **1978**, *17*, 1531–1548.
(42) Bowie, J.U.; Sauer, R.T. *Biochemistry* **1989**, *28*, 7139–7143.
(43) Huang, G.; Oas, T. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6878–6882.
(44) Piccoli, R.; Tamburrini, M.; Piccialli, G.; Donato, A.D.; Parente, A.; D'Alessio, G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 1870–1874.
(45) Plotkin, S.S.; Wang, J.; Wolynes, P.G. *J. Chem. Phys.* **1997**, *106*, 2392–2948.
(46) Chang, I.; Cieplak, M.; Dima, R.I.; Maritan, A.; Banavar, J.R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14350–14355.
(47) D'Aquino, J.A.; Friere, E.; Amzel, L.M. *Proteins:*

Struct., Funct., Genet. **2000**, *4*, 93–107.
(48) Papoian, G.A.; Ulander, J.; Eastwood, M.P.; Lutley-Schulten, Z.; Wolynes, P.G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3352–3357.
(49) Sobolev, V.; Wade, R.C.; Vriend, G.; Edelman, M. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 120–129.
(50) Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.B.; Meyer, E.F.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
(51) Pacios, L. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1427–1435.
(52) Koga, N.; Takada, S. *J. Mol. Biol.* **2001**, *313*, 171–180.
(53) Chavez, L.L.; Onuchic, J.N.; Clementi, C. *J. Am. Chem. Soc.*, in press.
(54) Clementi, C.; Nymeyer, H.; Onuchic, J.N. *J. Mol. Biol.* **2000**, *298*, 937–953.
(55) Clementi, C.; Jennings, P.A.; Onuchic, J.N. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5871–5876.
(56) Eastwood, M.P.; Wolynes, P.G. *J. Chem. Phys.* **2001**, *114*, 4702–4716.
(57) Kaya, H.; Chan, H.S. *J. Mol. Biol.* **2003**, *326*, 911–931.
(58) Robinson, C.R.; Sauer, R.T. *Biochemistry* **1996**, *35*, 13878–13884.
(59) D'Alessio, G. *Eur. J. Biochem.* **1999**, *266*, 699–708.
(60) Case, D.A.; Pearlman, D.A.; Caldwell, J.W.; Cheatham, T.E.; Ross, W.S.; Simmerling, C.L.; Darden, T.A.; Merz, K.M.; Stanton, R.V.; Cheng, A.L. et al. AMBER6, University of California, San Francisco, 1999.
(61) Ferrenberg, A.M.; Swendsen, R.H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
(62) Timm, D.E.; Haseth, P.L.d.; Neet, K.E. *Biochemistry* **1994**, *33*, 4667–4676.
(63) Tsai, C.-J.; Xu, D.; Nussinov, R. *Protein Sci.* **1997**, *6*, 1793–1805.
(64) Jones, S.; Thornton, J.M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13–20.
(65) Schreiber, G.; Fersht, A. *Nat. Struct. Biol.* **1996**, *3*, 427–431.
(66) Sheinerman, F.B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.
(67) Milla, M.E.; Brown, B.M.; Waldburger, C.D.; Sauer, R.T. *Biochemistry* **1995**, *34*, 13914–13919.
(68) Nolting, B.; Andert, K. *Proteins: Struct., Funct., Genet.* **2000**, *41*, 288–298.
(69) Becker, O.M. *J. Comput. Chem.* **1998**, *19*, 1255–1267.
(70) Levy, Y.; Caflisch, A. *J. Phys. Chem. B* **2003**, *107*, 3068–3079.
(71) Schlunegger, M.P.; Bennet, M.J.; Eisenberg, D. *Adv. Protein Chem.* **1997**, *50*, 61–122.
(72) Rousseau, F.; Schymkowitz, J.W.H.; Itzhaki, L.S. *Structure* **2003**, *11*, 243–251.
(73) Rousseau, F.; Schymkowitz, J.W.H.; Wilkinson, H.R.; Itzhaki, L.S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5596–5601.
(74) Schymkowitz, J.W.H.; Rousseau, F.; Wilkinson, H.R.; Friedler, A.; Itzhaki, L.S. *Nat. Struct. Biol.* **2001**, *8*, 888–892.
(75) D'Alessio, G. *Nat. Struct. Biol.* **1995**, *2*, 11–13.
(76) D'Alessio, G. *Prog. Biophy. Mol. Biol.* **1999**, *72*, 271–298.
(77) Oliveberg, M. *Acc. Chem. Res.* **1998**, *31*, 765–772.