

Garegin A. Papoian
Peter G. Wolynes
Department of Chemistry &
Biochemistry,
University of California at
San Diego,
9500 Gilman Dr.,
La Jolla, CA 92093-0371

Received 12 June 2002;
accepted 22 August 2002

The Physics and Bioinformatics of Binding and Folding—An Energy Landscape Perspective

Abstract: It has been recognized in the last few years that unstructured proteins play an important role in biological organisms, often participating in signal transduction, transcriptional regulation, and a variety of other regulatory activities. Various hypotheses have been put forward for the ubiquity of the unfolded state; rapid turnover, faster or more specific binding kinetics, multifunctionality may all possibly explain apparent ubiquitousness of unfolded proteins in eukaryotic cells. In this paper we extend the energy landscape theory of protein folding to construct an analytical model of how binding and folding are coupled thermodynamically when the energy landscape is partially rugged. To deduce the parameters that enter the theory, which is based on Generalized Random Energy Model, we have analyzed in a bioinformatic sense a large structural database of more than 500 protein complexes. We find that Miyazawa–Jernigan contact potential shows similar energy gaps for folding for both hydrophobic and hydrophilic proteins, but that for binding contacts hydrophobic interfaces turn out to be funneled while hydrophilic ones are antifunneled. This suggests evolution has found a mechanism for avoiding frustration between folding and binding by making use of indirect water-mediated interactions. By juxtaposing the monomeric protein folding free energy profile in the protein complex database with another database consisting of only well-folded monomers, we estimate that at least 15% of monomers in the former database are unfolded in the absence of partner protein interface interactions. When employing the parameters characteristic of these unfolded monomers to construct binding/folding phase diagrams, we find that these monomers would indeed fold if sufficiently stabilizing binding contacts, consistent with that fold, are formed. © 2003 Wiley Periodicals, Inc. *Biopolymers* 68: 333–349, 2003

Keywords: bioinformatics; binding; folding; energy landscape theory; Generalized Random Energy Model; Miyazawa–Jernigan contact potential; binding-folding funnel

INTRODUCTION

Protein–protein interactions play the role of the nervous system in the cell, orchestrating and guiding such vital processes as cell proliferation, differentia-

tion, and death.¹ Given the paramount importance of protein–protein interactions in biology and health sciences, the elucidation of the mechanism of these interactions has attracted enormous research effort in the last century or so. The celebrated “lock-and-key”

Correspondence to: Peter G. Wolynes; email: pwolynes@ucsd.edu

Contract grant sponsor: National Institute of Health
Biopolymers, Vol. 68, 333–349 (2003)
© 2003 Wiley Periodicals, Inc.

paradigm of protein association^{2,3} may be rightly considered as one of the key bedrocks of modern molecular biology. Yet another model for protein–protein interactions, the so-called induced fit mechanism,⁴ has also helped to provide immensely useful insights into biological processes where conformational plasticity plays an important role. In this paper we begin to build a basic analytical theory for a third mechanism of protein association, in which large unstructured pieces of protein chain order uniquely only when binding is initiated.

The lock-and-key mechanism of protein binding has been viewed by the bulk of biological scientists as the universal way proteins interact with each other. In the lock-and-key description two proteins are imagined as being rigid bodies with sterically and electrostatically complementary surfaces that lock into each other upon interaction. Important biological functions, such as enzyme catalysis, immunological recognition, or molecular discrimination, require very fine control of the protein's three-dimensional structure, thus these proteins are expected to have a rigid folded structure in their active state.^{5,6} The theoretical description of a lock-and-key process can be captured by docking protocols in which two rigid proteins or a rigid protein and a ligand are brought together and screened using shape- or energy-based methods.⁷ Such protocols are a mainstream of a modern drug design.^{8,9} As for the calculation of the kinetic rate constants of protein association, it has been recognized that a naive elastic collision model based on estimating the rate of “lucky” encounters when the “key” and the “lock” are perfectly aligned for the interaction to succeed produces many orders of magnitude smaller rate constants than experimentally observed. To explain this nonspecific adhesive encounters along with rotational diffusion, electrostatic steering, and guidance by coupling to the funneled landscape have been put forward as three routes to the observed, rate constants that have values close to the diffusion limit for small molecules.^{10–12}

The more complex induced fit mechanism was first invoked to explain the allosteric control of proteins by ligands, where an alternative, less favorable protein conformation in equilibrium with the native state interacts preferentially with the given ligand, which in turn shifts correspondingly the conformational equilibrium.¹ The equilibrium between tense T macrostate and relaxed R macrostate is influenced by ligand binding, which in turn leads to concerted or sequential conformation changes in a multisubunit protein.¹ The shifting of the conformational equilibrium toward a less populated state due to ligand binding has been

recently explored within the context of the energy landscape model.^{13–15}

Yet the deviations from the lock-and-key picture may be more profound. It has become increasingly apparent in the last decade or so that a large number of interacting proteins, especially those implicated in eukaryotic cell differentiation and signaling, are either very flexible or completely unfolded in the free state, implying the simultaneous occurrence of binding and folding.^{5,6,16–22} Although it might seem somewhat counterintuitive that fully functional important proteins in the working cell would exist as random coils or molten globules, Wright and Dyson have recently provided several powerful arguments that may explain the role of unfolded proteins in the cell.^{5,6} For instance, they argued, cell cycle, transcriptional, and translational proteins will be targeted for rapid turnover by being unfolded, thus providing an additional lever of control. Promiscuous binding, allowing nonlinear control of many processes, and in a contrary sense, higher specificity have both been suggested as additional reasons for having highly floppy proteins in the cell.⁶ As for the ubiquity of at least partially (>50 residues) unfolded monomers in biological organisms, over 100 unique functioning proteins having more than 250 homologs have been experimentally verified as being partially unstructured.²³ Neural network studies by Dunker and co-workers have suggested that up to 30% of sequences of eukaryotic genomes may correspond to proteins containing large disordered segments.²⁴

Since the importance of highly unstructured proteins has been recognized only recently, only a few theoretical papers have begun to address this issue and its implications. Shoemaker, Portman, and Wolynes have investigated the kinetic aspects of simultaneous binding and folding, concluding that being unfolded leads to somewhat faster rate of binding due to protein's greater capture radius (the so-called “fly-casting” mechanism).²⁵ Terada, Sasai, and Yomo have applied the analytical techniques of energy landscape theory to study the actin–myosin system, suggesting that the ATP hydrolysis energy is channeled into local unfolding of myosin head, which in turns drives the unidirectional motion of actomyosin molecular motor.²⁶

In contrast to these papers we focus here on the thermodynamics of folding/binding while accounting for the complexity of the protein energy landscape. The energy landscape picture of protein folding, which has proven to be very useful in rationalizing many details of the folding phenomenon,^{13,27,28} serves as the starting point to derive this connection from the statistics of interresidue interactions and

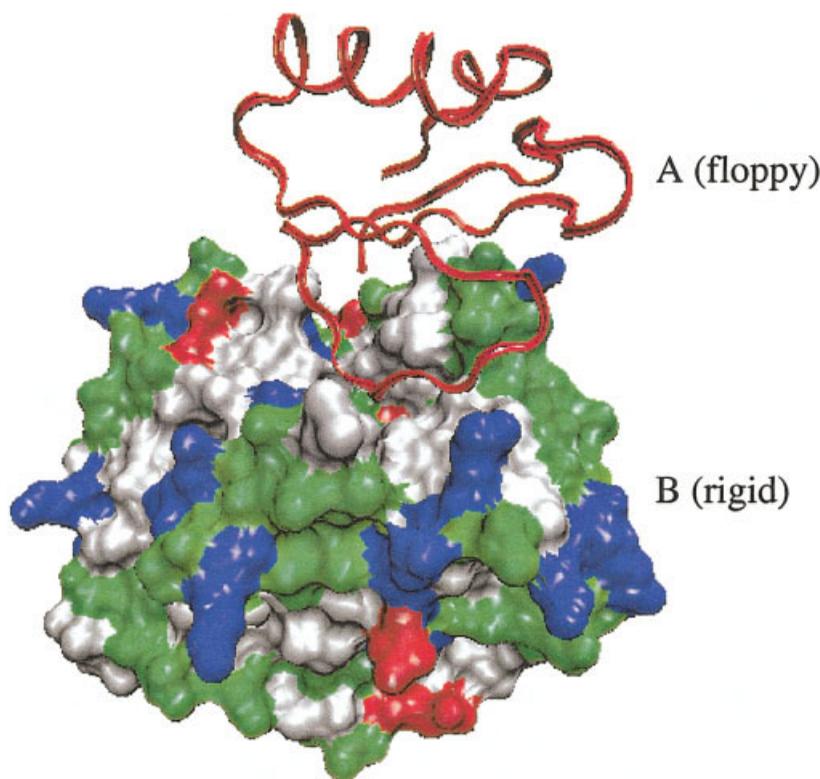


FIGURE 1 The schematic representation of floppy protein A interacting with the surface of rigid template protein B.

polymer physics. It also is a natural approach to the question of specificity and binding promiscuity, which we hope to address in future papers. We obtain a free energy expression for binding–folding. This theory contains several statistical parameters that we infer by bioinformatically analyzing a structural database of more than 500 carefully chosen protein complexes compiled by Ben-Tal and co-workers.²⁹ By comparing the folding free energy profile between the protein complex database and another single monomer database, we estimate the smallest fraction of monomers that are expected to be unfolded in the former database. For the model parameters characteristic of unfolded proteins, we show that native binding interactions do indeed lead to simultaneous folding as well.

BASIC THEORY

To set up the model we partition the system of interacting proteins into two major thermodynamic states—the associated complex and the dissociated monomers. The former complex may be divided further into several additional substates, which differ by

the degree of folding of partner proteins as well as by the degree of ordering of the interface contacts. The transition between these various *associated* states is the major goal of our study; the thermodynamics of the dissociated state may be analyzed in a straightforward manner in the limit of low concentrations.

According to our model, unfolded protein A forms a transient collision complex with rigid protein B, as schematically depicted in Figure 1. During the initial phase of the encounter complex formation, protein A stays unfolded and forms *random* binding contacts with the surface of B. If the energy landscape for binding is strongly funneled and is *correlated* with the folding funnel, then the complex undergoes a binding/folding phase transition. Otherwise, if in addition the nonnative binding interactions are not strong enough, the complex dissociates.

In order to quantify the description of the sequence of events given above, we introduce two order parameters, Q_f and Q_b , which indicate the degree of similarity between the set of *folding* and *binding* contacts in the given configuration and the native folded/bound reference state. These order parameters vary from zero in the completely nonnative state to one in the reference native state. Our goal is to develop thermo-

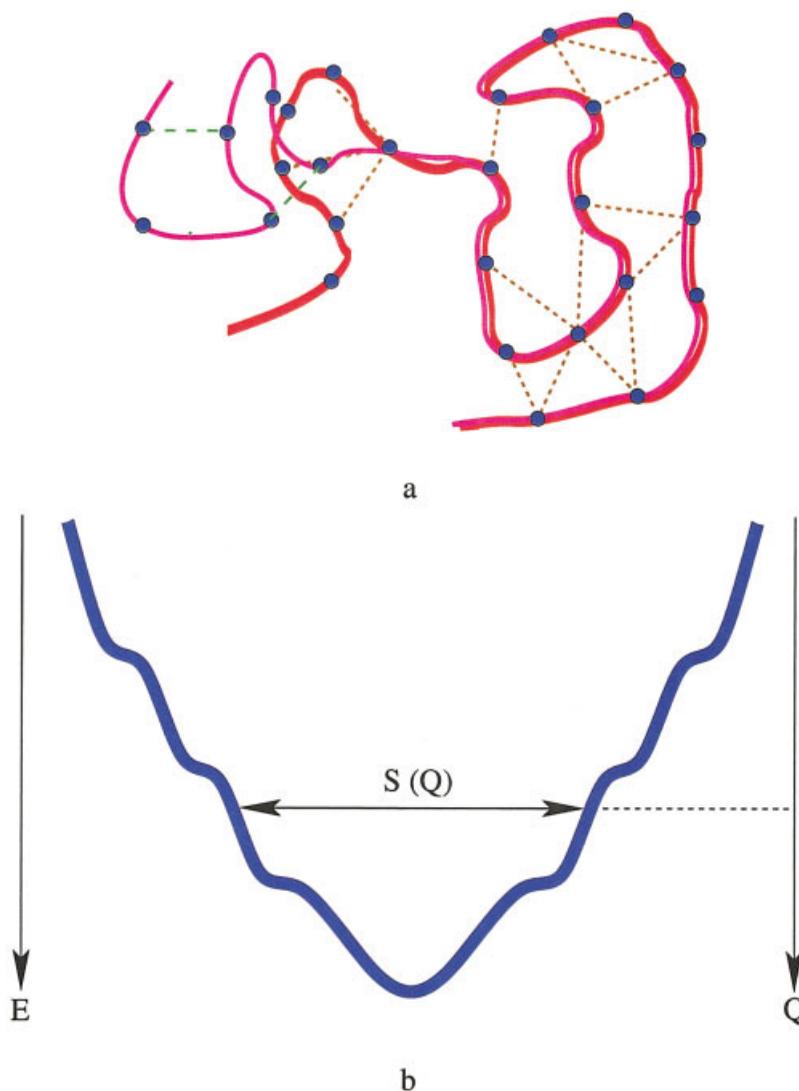


FIGURE 2 Basic premises of the stratum GREM formalism. (a) A given structure (in magenta) has a number of overlapping contacts with the native structure (red); (b) energy funnel as a function of overlap parameter Q .

dynamic formalism that will allow us to calculate free energy as a function of these two order parameters, which in turn might indicate the necessary conditions for folding and/or binding phase transitions.

Derivation of the Free Energy Expression

The Generalized Random Energy Model (GREM) of protein folding serves as a convenient starting point for the development of our binding/folding coupling model. According to the GREM picture of protein folding, the multitude of protein conformations is partitioned into strata in such a way that the conformations having the same similarity measure with the

unique native state are grouped together³⁰ (Figure 2). These ensembles of states are then treated statistically within the context of the Random Energy Model (REM) formalism, as having a particular energy mean and the variance. As the similarity measure Q is increased from 0 to 1, more narrow strata are cut on the energy folding funnel (see Figure 2), leading to conformational ensembles that are low in energy, have smaller energy variances as well as smaller numbers of available states (i.e., lower entropy). The interplay between the lower mean energy on one hand and the lower entropy and the energy variance on the hand determines the shape of the protein folding free energy landscape. Plotkin, Wang, and Wolynes have

derived the following expression for the free energy of this model:³⁰

$$F(Q) = zN\delta E_f Q - NS^0(Q)T - \frac{zN(1 - Q^2)\Delta E^2}{2k_B T} \quad (1)$$

where N is the protein chain length, z is the number of contacts per residue, δE_f is the energy gap per residue between the native state energy, and the average energy of the denaturated ensemble, ΔE^2 , is the energy variance per residue of the denaturated ensemble, s^0 is the conformational entropy per residue, and T is the temperature.

In order to derive a free energy expression equivalent to Eq. (1) for the case where binding and folding contacts occur simultaneously (Figure 1), we start from the following microscopic statistical contact Hamiltonian:

$$H = H_f + H_b \quad (2)$$

$$H_f = \sum_{i < j} (\Delta_{ij}^f \sigma_{ij}^f \epsilon_{ij}^f + (1 - \Delta_{ij}^f) \sigma_{ij}^f \epsilon_{ij}^{nn}) \quad (3)$$

$$H_b = \sum_{i \forall A, j \forall B} (\Delta_{ij}^b \sigma_{ij}^b \epsilon_{ij}^b + (1 - \Delta_{ij}^b) \sigma_{ij}^b \epsilon_{ij}^{nn}), \quad (4)$$

where H_f and H_b are folding and binding Hamiltonians respectively; Δ_{ij} is 1 if i and j are in contact in the *native* conformation, 0 otherwise; σ_{ij} is 1 if i and j are in contact in the *given* conformation “a,” 0 otherwise; $\epsilon_{ij}^{b/f}$ and ϵ_{ij}^{nn} are random energy variables for native and nonnative contacts respectively (f/b subscripts/superscripts refer to binding and folding). The factor $\sum_{i,j} \Delta_{ij}^{b/f} \sigma_{ij}^{b/f}$ is the total number of contacts shared between conformation “a” and the binding/folding native conformations, while $\sum_{i,j} (1 - \Delta_{ij}^{b/f}) \sigma_{ij}^{b/f}$ is the number of contacts unique to “a.” Since a certain number of contacts are shared between “a” and the native binding/folding conformations, those shared parts of Hamiltonian will lead to correlations in the energy of two structures. Thus, the total energy of the native folding and binding states as well as conformation “a” may be partitioned in the following way:

$$E_f^0 = E_{af} + E_f' \quad (5)$$

$$E_b^0 = E_{ab} + E_b' \quad (6)$$

$$E_a^0 = E_{af} + E_{ab} + E_a'. \quad (7)$$

Those contacts that are shared between “a” and the native folding and binding contacts lead to common

energy terms E_{af} and E_{ab} , while contacts unique to each conformation are denoted by primes. Since the energy terms on the right side of Eqs. (5)–(7) are random variables comprised from the respective sums of the microscopic contact energy random variables (ϵ 's), the simultaneous occurrence of E_{af}/E_{ab} in more than one equation couples the resulting conformation energies (i.e., E_f^0 , E_b^0 , and E_a^0). If we require that native folding and binding energies E_f^0 and E_b^0 adopt certain (low energy) values, then the conditional probability of finding E_a^0 at a particular energy may be found from

$$P(E_a^0 | E_f^0, E_b^0) = \frac{\langle \delta[E_a^0 - H(Q_f, Q_b)] \delta[E_f^0 - H_f(Q_f)] \times \delta[E_b^0 - H_b(Q_b)] \rangle_{E_{af}, E_{ab}, E_f', E_b'}}{\langle \delta(E_f^0 - H_f(Q_f)) \rangle_{E_f'} \langle \delta(E_b^0 - H_b(Q_b)) \rangle_{E_b'}}, \quad (8)$$

where δ denotes the Delta function, and angular bracket subscripts indicate averaging with respect to the corresponding random energy variables. If $\Omega(Q_f, Q_b)$ is the total number of available states, then the microcanonical entropy may be written as

$$S(E_a^0, Q_f, Q_b) = k_B \ln(\Omega(Q_f, Q_b) P(E_a^0 | E_f^0, E_b^0)) \quad (9)$$

The temperature dependence is introduced with the help of a standard relation, $\partial S(E_a^0) / \partial E_a^0 = 1/T$, leading to the following expression for free energy:

$$F(Q_f, Q_b) = z_f N \delta E_f Q_f + z_b N \delta E_b Q_b - NS^0(Q_f, Q_b)T - \frac{z_f N (1 - Q_f) (1 + \gamma_f Q_f) \Delta_{\epsilon_f}^2}{2k_B T} - \frac{z_b N (1 - Q_b) (1 + \gamma_b Q_b) \Delta_{\epsilon_b}^2}{2k_B T} \quad (10)$$

where γ is the degree of native contact heterogeneity (i.e., the ratio of $\Delta \epsilon_{\text{native}}^2$ to $\Delta \epsilon_{\text{nonnative}}^2$) and δE_f and δE_b are the gaps between the energies of the native folding and binding states and the average energy of the nonnative states ensemble. A quick analysis of the free energy expression given above reveals that the folding and binding coordinates become intimately coupled through the nonadditive entropy term, $S^0(Q_f, Q_b)T$. In addition, several unknown parameters entering Eq. (10) need to be determined in order to construct corresponding phase diagrams. The derivation of an entropy formula as well as the parameterization of Eq. (10) is the subject of the next two sections.

Entropy Model

One has a number of choices of various degree of sophistication to set up a model of entropy for a protein chain as a function of some average order parameters. For instance, for the case of a single protein folding, Plotkin, Wang, and Wolynes have used the similarity measure Q between the given structure and the native structure as the reaction coordinate, deriving a polymer physics based entropy formulas for low and high Q limits and interpolating in between.³¹ In the current context of binding and folding, a similar procedure, although possible, would become excessively cumbersome. This prompted us to use a simpler quasi-linear model, since the earlier computer simulations of simplified polymer chains suggested the near linearity of entropy as a function of Q in the context of protein folding.^{32,33} Our present approach is to calculate entropy for four corners of the $Q_f = 0 \dots 1$, $Q_b = 0 \dots 1$ domain and interpolate in between with a bilinear $\{Q_f, Q_b\}$ form.

The $Q_f = 1$, $Q_b = 1$ vertex corresponds to the completely ordered bound/folded state, thus its entropy, $S(1, 1)$, serves as a reference point, taken to be zero for convenience. The $Q_f = 1$, $Q_b = 0$ corner represents the natively folded protein which undergoes solid-body rotational-translational diffusion, sampling the surface of the partner protein. The entropy for this state, $S(0, 1)$, is calculated below. The $Q_f = 0$, $Q_b = 1$ state corresponds to ordering of interface contacts, while the rest of the protein being unfolded, and we calculate its entropy $S(1, 0)$ using polymer physics techniques. Finally, the $Q_f = 0$, $Q_b = 0$ vertex corresponds to an unfolded protein executing a rotational-translational diffusion near the surface of the partner protein, and we estimate the corresponding entropy $S(0, 0)$ as a sum of the entropy of the unfolded free chain and the abovementioned rotational-translational entropy $S(1, 0)$.

Given the entropy values for the four corners of the Q_f, Q_b quadrant, the bilinear formula in Eq. (11) incorporates those into a smooth (quasilinear) surface, interpolating in between:

$$S^0(Q_f, Q_b) = (1 - Q_f)(1 - Q_b)S(0, 0) + Q_b(1 - Q_f)S(0, 1) + Q_f(1 - Q_b)S(1, 0) \quad (11)$$

For the entropy of folding of the isolated free protein, we use the experimental estimate of $\approx 3k_B$ per residue by Freire and co-workers.^{34–36} Since $S(0, 0)$ is the sum of this entropy and $S(1, 0)$, we proceed next to estimate the latter quantity.

In the absence of native binding contacts, the protein is free to explore the surface of the partner pro-

tein. To be considered as part of the associated state, the center of mass of the protein, which is idealized as being spherical, must remain within the region of space enclosed by two concentric spheres, as depicted in Figure 3. The thickness of this layer is determined by the contact cutoff, 6.5 Å in our model. As for the completely ordered reference state, $Q_f = 1$ and $Q_b = 1$, the protein's center of mass is restricted to move only within the region bound by the intersection of two concentric sphere mentioned above and a cone with solid angle $\theta = \beta$ and $\phi = \beta$. The volume enclosed in that region is indicative of the size of the native basin for the center of mass of the completely ordered bound and folded protein. Solid angle β is determined by the intersection of a contact sphere centered on surface of the central protein with protein's surface (see Figure 3). The reduction in translational entropy when native binding contacts are formed, due to the additional confinement of protein's center of mass, is computed as

$$\Delta S_{\text{transl}} = k_B \ln \left(\frac{1}{9} \frac{d^2}{R^2} \right) \quad (12)$$

where d indicates contact radius and R indicates protein's radius. For the average radius of 19.3 Å in the protein complex database, we have calculated 4.4 k_B loss of translational entropy upon ordering of interface contacts.

When no native binding contacts are formed, the protein is free to diffuse rotationally around its principal axes. The rotational entropy associated with such motion is lost to a certain degree, when the native interface forms. In the latter case, the rotational motion of the protein is confined to an angle comparable to β around each of the three principal axes (Figure 3). By computing the rotational partition functions for free rotation and for constrained rotation, the reduction in rotational entropy due to native interface formation may be calculated as

$$\Delta S_{\text{rot}} = k_B \ln \frac{\beta^2 [1 - \cos(\beta)]}{8\pi^2} \quad (13)$$

where β was defined above. For the average protein radius of 19.3 Å and contact radius of 6.5 Å, we estimate 5.3 k_B loss in solid-body rotational entropy due to confinement upon native binding contact formation. Thus, the combined entropy loss of 9.7 k_B defines the value for $S(1, 0)$. In addition, it has to be added to 3 k_B per residue estimate of free chain entropy of folding to obtain $S(0, 0)$. Notice, however, that for a 200 residue protein this amounts only to

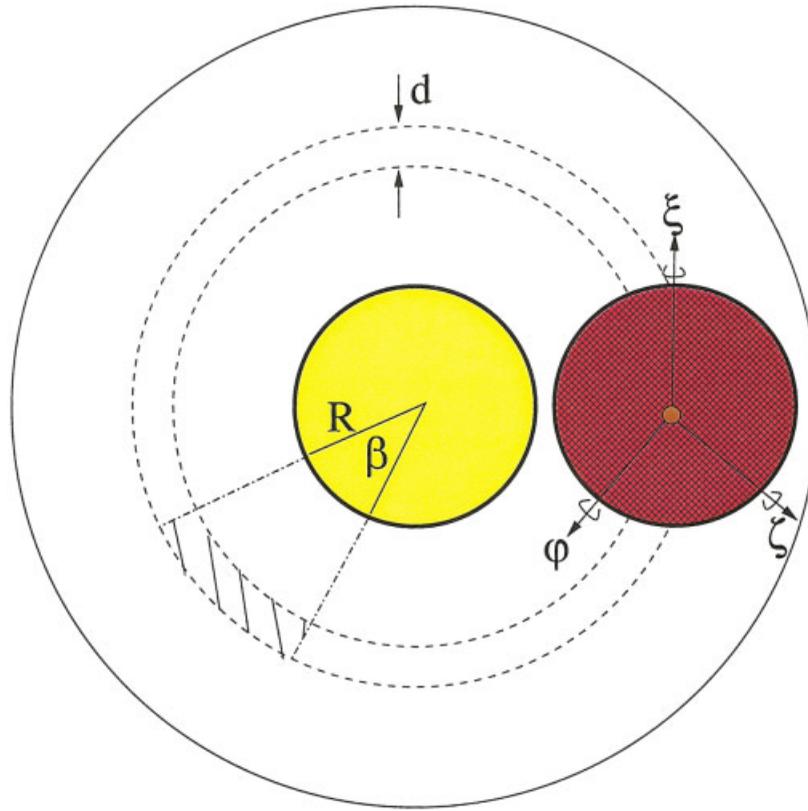


FIGURE 3 Entropy of the folded protein, which moves diffusively on the surface of the partner protein is reduced when native binding interface is formed. The dashed concentric circles indicate the region of space where the protein center of mass is allowed to explore in order to be considered as part of the associated state.

additional $0.05 k_B$ entropy per residue, i.e., very small compared with the folding entropy.

The last vertex to be determined is $Q_f = 0$ and $Q_b = 1$, in which case interface contacts are formed in full but the protein is still unstructured internally. The entropy of folding for free chain, S_f^0 , is reduced by a certain factor if the protein forms native contacts on the binding partner protein's surface (Figure 4). This reduction of entropy may be attributed to the following three sources: (1) residues directly in contact with surface lose all free chain entropy, (2) there is a reduction in entropy due to end segments, (3) there is a reduction in entropy due to surface loop formation. By analyzing structurally the database of 500+ protein-protein complexes mentioned above, we have obtained the ratios of the number of residues in sequentially in contact, the number in the surface end segments, and the number in surface loops as 0.13:0.40:0.47. In addition, we have extracted from the structural database the frequency of end segment and surface loop formation as a function of chain length. Thus, if the reduction in entropy were to be known for end segments and surface loops as a function of

number of residues involved, the total reduction in entropy [i.e., $S^0(Q_f = 0, Q_b = 1)$] may be easily computed by using the database derived structural partitioning coefficients to do the summations.

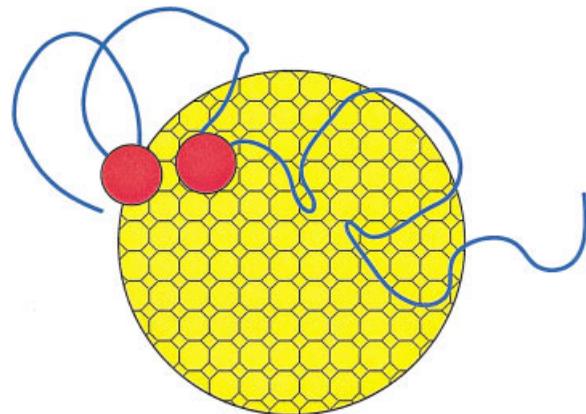


FIGURE 4 Entropy of the free chain is reduced when forming loops and sequential trains of residues on the partner proteins surface. Small red spheres indicate the contact volumes.

The reduction in free chain entropy for end segments and surface loops is caused by conformational constraint of having to avoid an obstacle, the “sphere” occupied by partner protein (4). For the limit of long chains, the probability of starting a chain at point \mathbf{R} and ending at point \mathbf{R}' is given by the Green function solution $[G(\mathbf{R}, \mathbf{R}'; N)]$ to the diffusion equation with absorbing boundary conditions at sphere’s surface:³⁷

$$\left(\frac{\partial}{\partial N} - \frac{b^2}{6} \nabla^2\right) G(\mathbf{R}, \mathbf{R}'; N) = \delta(\mathbf{R} - \mathbf{R}') \delta(N) \quad (14)$$

$$G(\mathbf{R}, \mathbf{R}'; N) = 0, \|\mathbf{R}'\| = R_{\text{sph}} \quad (15)$$

where b is the chain persistence length, taken to be 20 Å as in the earlier works.³⁸ The solution to Eqs. (14) and (15) is given by³⁹

$$G(\mathbf{R}, \mathbf{R}'; N) = \frac{1}{\sqrt{RR'}} \sum_{n=0}^{\infty} (2n+1) P_n(\mu) \times \int_0^{\infty} \frac{C_{n+(1/2)}(uR) C_{n+(1/2)}(uR')}{J_{n+(1/2)}^2(uR_{\text{sph}}) Y_{n+(1/2)}^2(uR_{\text{sph}})} e^{-(b^2/6)u^2 N} u du \quad (16)$$

$$C_{n+(1/2)}(z) = J_{n+(1/2)}(z) Y_{n+(1/2)}(uR_{\text{sph}}) - Y_{n+(1/2)}(z) J_{n+(1/2)}(uR_{\text{sph}}) \quad (17)$$

where $P_n(\mu)$ is a Legendre polynomial with $\mu = \cos \theta$, θ being the angle between \mathbf{R} and \mathbf{R}' ; $J_{n+(1/2)}(z)$ and $Y_{n+(1/2)}(z)$ are spherical Bessel functions. The Green function, as given by Eq. (16), may be used to estimate the probability of end segment and surface loop formation:

$$P_{\text{loop}}(\theta; N) = \frac{1}{V_{\Gamma}} \int_{\mathbf{R} \in \Gamma} \int_{\mathbf{R}' \in \Gamma'} G(\mathbf{R}, \mathbf{R}'; N) d\mathbf{R} d\mathbf{R}' \quad (18)$$

$$P_{\text{end}}(N) = \frac{1}{V_{\Gamma}} \int_{\mathbf{R} \in \Gamma} \int_{\|\mathbf{R}'\| > R_{\text{sph}}} G(\mathbf{R}, \mathbf{R}'; N) d\mathbf{R} d\mathbf{R}' \quad (19)$$

where Γ and Γ' refer to the regions of origination and destination contacts (red spheres in Figure 4) and V_{Γ} is the contact volume.

We have surveyed the protein complex database for the typical values of R_{sph} and θ parameters entering Eqs. (18) and (19). The average protein radius was found to be 19.3 ± 4.5 Å, while the average angle of loop formation was observed to be 0.13 ± 0.02 ra-

dian. By additionally taking into account the ratios of trains, end segments, and loops of 0.13:0.40:0.47, as well as the corresponding distributions of end segments and loops vs chain length (not shown), we have calculated the relative retention in entropy as a function of overall chain length (Figure 5a). The range and the shape of the curve in Figure 5a is not significantly dependent on the variation of the partner protein’s radius within the standard deviation of the database as well as the angular distribution of loop end points. On the other hand, if the initial estimate of the folding entropy per residue is lowered from $3 k_{\text{B}}$ to $1.5 k_{\text{B}}$, then the ratio of S^0 ($Q_{\text{f}} = 0$, $Q_{\text{b}} = 1$) to S^0 ($Q_{\text{f}} = 1$, $Q_{\text{b}} = 0$) changes from 0.77 to 0.62 for a 200-residue protein, indicating moderate sensitivity to the estimate of the absolute entropy of folding.

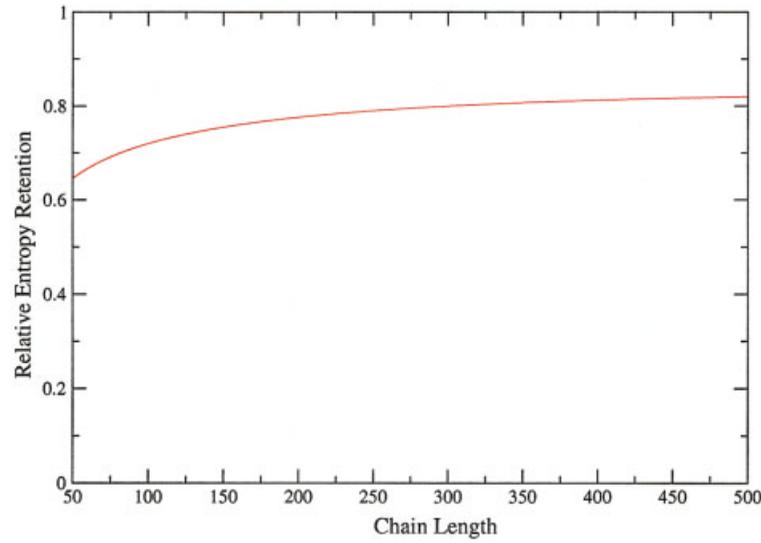
The dependence of the binding/folding entropy as a function of Q_{f} and Q_{b} is plotted in Figure 5b by substituting the value for S^0 ($Q_{\text{f}} = 0$, $Q_{\text{b}} = 1$) for a 200-residue protein read from Figure 5a into Eq. (11). As we have alluded earlier, the shape of the entropy surface is nearly linear.

DATA MINING: EXTRACTING THE GAP AND THE VARIANCE PARAMETERS

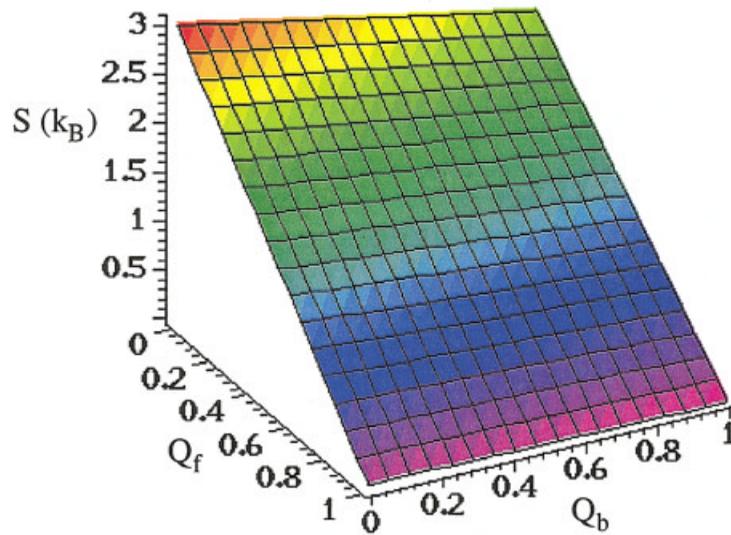
One of the main goals of our current study was to survey a realistic database of protein–protein complexes for the variability of gaps and variances so we could approximately estimate the proportion of proteins that are unfolded when isolated. Another intriguing question is the relative depth of the binding funnel compared with the folding funnel. To carry out our study we have chosen, as indicated earlier, a non-redundant database of more than 500 protein–protein complexes published by Ben-Tal and co-workers.²⁹

To calculate energies of proteins, we have used the well-known Miyazawa–Jernigan (MJ) pairwise additive contact potential.⁴⁰ A few other pairwise contact potentials have been recently published in the literature;^{41–44} however, they are often found to correlate strongly with the Miyazawa–Jernigan potential.

Although the calculation of the energies of the *native* structures as found in the database is rather simple with the MJ potential, the modeling of the denaturated state is less straightforward. One option, albeit very costly, is to carry out a long molecular dynamics or Monte Carlo simulation of each protein individually and in complex with its corresponding partner (≈ 1500 such simulations). An alternative approach, the one that we chose for the current work, is to shuffle the protein sequence to imitate alternative conformations. Similar permuting the protein contact



a)



b)

FIGURE 5 (a) The ratio of retained entropy on forming binding only native contacts to free entropy as a function of chain length. (b) Entropy as a function of folding and binding order parameters for a 200-residue protein.

maps has been employed by Domany and co-workers to study protein folding.⁴⁵ Some of the contact maps introduced by sequence shuffling are presumed to be nonphysical due to the chain connectivity constraint; however, there exist in principle algorithms for examining the viability of the resulting contact maps.⁴⁶ In addition, these errors are expected to be less important since we are not interested in protein chain dynamics but only in average properties. The sequence shuffling averages and the real chain dynamics averages are believed to converge in the high temperature limit.

The following protocol was employed for generating the partially (Q) denaturated ensemble sequences (“conformations”). A random residue in contact was permuted with another randomly chosen residue and the number of native contacts affected by such move was recorded. This step was repeated the necessary number of times until the number of altered contacts reached Q time the total number of contacts. After calculating the energy of the resulting sequence conformation, the whole procedure was repeated 10000 times for each protein individually as well as for

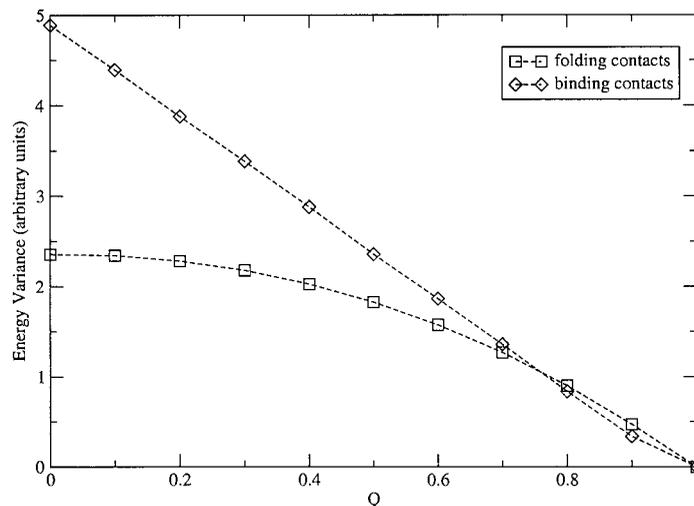


FIGURE 6 The energy variances per contact of the folding and binding contacts as the functions of similarity measure Q to the native state.

protein–protein interfaces. The resulting energy variance per contact as a function of Q is shown in Figures 6.

We have showed earlier [Eq. (10)] that the folding and binding variances vary with Q as $(1 - Q)(1 + \gamma Q)\Delta E^2$, where γ indicates the degree of heterogeneity of native contacts ($0 \leq \gamma \leq 1$). When this formula is fit into Figure 6, the folding native contacts are found to be as heterogeneous as nonnative contacts ($\gamma_f \approx 1$), while native binding contacts are much more homogeneous ($\gamma_b \approx 0$).

Our calculations of the energy gap as a function of Q for folding and binding contacts has revealed an

interesting trend. We partitioned both folding and binding contacts into nearly equal groups according to their respective hydrophobicity composition (hydrophobicity scale was taken from Ref. 47). Folding contact curves indicate that the Miyazawa–Jernigan potential produces nearly indistinguishable energy gaps for both hydrophilic and hydrophobic compositions (Figure 7). The binding contacts, on the other hand, present striking dissimilarity—hydrophobic interfaces are funneled while hydrophilic ones are antifunneled (Figure 7). These results suggest a qualitative difference in the forces involved in the two different classes of binding interfaces. Very clearly,

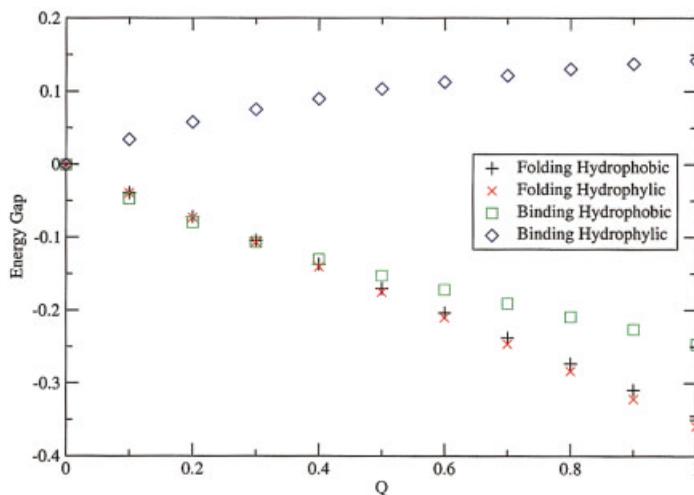


FIGURE 7 The energy gaps per contact of the folding and binding contacts as the functions of similarity measure Q to the native state. The proteins are partitioned into being either hydrophobic or hydrophilic.

specific water-mediated interactions must play an important role in hydrophilic interfaces. In essence, it appears evolution has come up with a qualitatively new strategy for overcoming what would otherwise be a strong conflict (i.e., frustration) between the requirement of folding and binding. Thus we see a new bioinformatic route to finding a water mediated contact potential that will eliminate the frustration too.⁴⁸

We have sidestepped so far the issue of the absolute energy scale characteristic to the Miyazawa–Jernigan potential. Since no conclusive arguments were provided for the particular value of the Miyazawa–Jernigan energy scale, we have chosen to infer it from the folding REM model. When the free energy of the folded state ($Q = 1$) for an isolated protein is equated to the free energy of the denatured ensemble ($Q = 0$) at the folding transition temperature T_F , then Eqs. (20)–(22) result from Eq. (1):

$$\delta E_f = -S^0 T_F - \frac{\Delta E^2}{2k_B T_F} \quad (20)$$

$$\delta E_f = zX \quad (21)$$

$$\sqrt{\Delta E^2} = b \sqrt{zX} \quad (22)$$

where z is the number of contacts per residue, b is inversely related to the energy gap over the square root of the variance ratio, and X indicates the yet to be determined energy scale. When definitions (21)–(22) are substituted into Eq. (20) and the consequent equation is solved for X , the following equations relate X to b :

$$X = T_F \frac{-2zk_B + 2\sqrt{z^2 k_B^2 - 2zS^0 k_B b^2}}{2zb^2} \quad (23)$$

$$b_{\max} = \frac{\sqrt{2zk_B S^0}}{2S^0} \quad (24)$$

where the maximum value of b , b_{\max} , is given by the requirement of having a non-negative value under the square root expression in Eq. (23). The upper graph on Figure 8 depicts the energy scale X (in units of $k_B T_F$) as a function of b , when substituting the estimate^{34–36} of $3k_B$ for the entropy per residue (S^0) in the unfolded ensemble as well as 340 K as the average folding transition temperature into Eq. (20). The corresponding dependence of the glass transition temperature of the underlying REM model is shown in the lower graph of Figure 8. At the maximum value of 0.55 for b the unfolded protein glass transition tem-

perature approaches the folding transition temperature. It has been argued before that the ratio of T_F over T_g is near 1.6 for natural proteins, which corresponds to the crossing of dashed lines in Figure 8 ($b = 0.5$, $X = -2.3k_B T_F$).

Given a certain value for entropy S^0 , there exist only a narrow range of b values that satisfy the folding REM model. Alternative estimates of entropy may be made if there is an assumption of significant secondary structure formation in the denatured state;²⁷ however, the b range remains rather limited. For instance, if the entropy is estimated at much lower $0.6 k_B$, b_{\max} becomes only 1.1, still lower than 3.1 as directly calculated with the Miyazawa–Jernigan contact potential. Thus, taking into account the apparent glassiness of the Miyazawa–Jernigan potential, we have chosen to scale uniformly calculated variances by certain factors for folding and binding contacts respectively so to conform to the $b = 0.5$ requirement (i.e., $T_F/T_g = 1.6$). Since we are interested in gauging the parameter space variability among protein–protein complexes in the databank, the rescaling of binding and folding variances by appropriate constant factors is expected not to affect our analysis.

RESULTS AND DISCUSSION

Having determined parameters for Eq. (10), we are ready to calculate phase diagrams for coupling of binding and folding for database protein complexes. Before doing this, however, it is appropriate to estimate the fraction of individual monomers in the database that are unfolded in the absence of their partner proteins. If this fraction is exceedingly low, then the monomers are completely folded on their own and the issue of coupling binding with folding becomes a moot one.

We start our analysis by calculating free energies of folding for all individual monomers in the database by Ben-Tal and co-workers (called Protein Complex Database from this point on). Histograms of the fractions of proteins having a free energy of folding in a particular window are shown in Figure 9. From the integration of the solid curve in Figure 9a we infer that 31% of monomers in the database have larger than +1 kJ/residue free energy of folding at 300 K (i.e., these monomers are unfolded at room temperature). Since it is hard to estimate the absolute accuracy of the Miyazawa–Jernigan potential as well as the Random Energy Model used, a comparison of these results with some additional reference database may provide further insights.

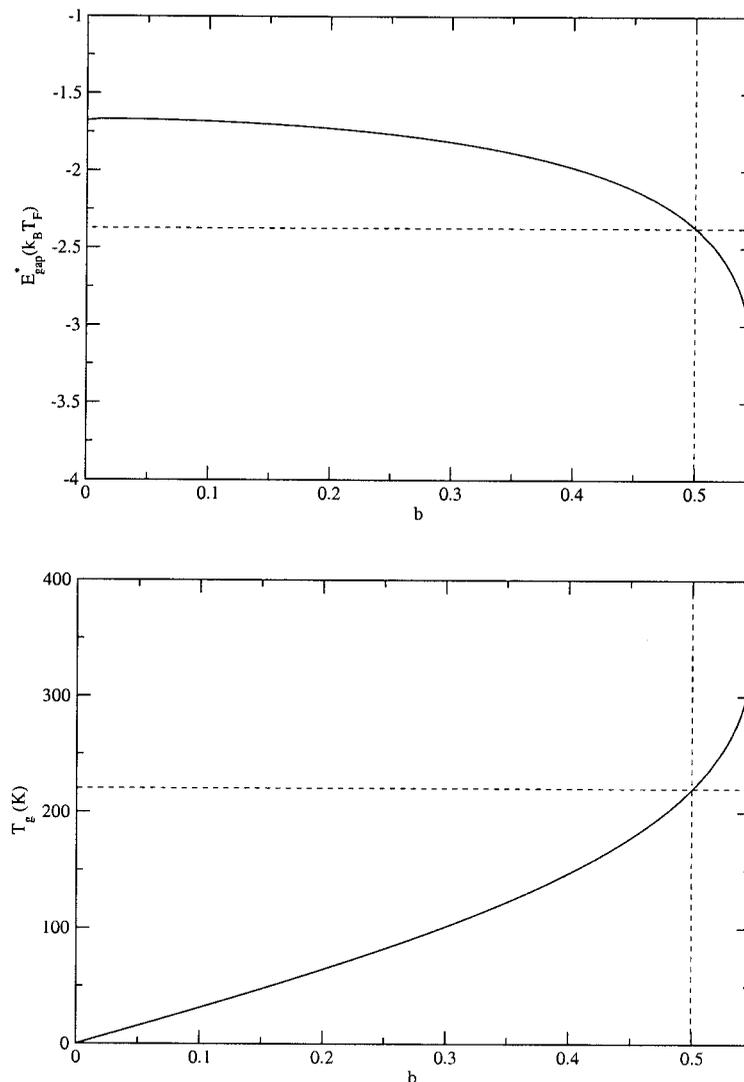


FIGURE 8 Upper graph: The energy gap ($\delta E_f^* = \delta E_f / z_f$) of the folding REM model in absolute energy units $k_B T_F$ ($T_F = 340$ K) as a function of $b = \delta E_f / \sqrt{\Delta E^2} \sqrt{z_f}$. Lower graph: The glass transition temperature of the underlying REM model as a function of b .

Banavar and co-workers have compiled a database (called Single Protein Database from this point on) that includes exclusively proteins that were crystallized as monomers.⁴⁹ If one makes an assumption that all proteins in that database spontaneously fold at room temperature, then the calculated database folding profile (dashed curves in Figure 9) may be used to calibrate the corresponding Protein Complex Database profile. The comparison of solid and dashed curves in Figure 9 suggests that indeed a larger fraction of individual monomers are unfolded in the Protein Complex Database than in the Single Protein Database. In particular, 16% of proteins in the Single Protein Database are found to have larger than +1 kJ/residue free energy of folding, thus prompting one

to conclude that at least on the order of $31 - 16 = 15\%$ of monomeric chains in the Protein Complex Database are unfolded while isolated.

To gain further insight into the nature of possibly unfolded monomers in the Protein Complex Database, we have computed (see Table I) the 20 most unstable monomers in the database. A quick perusal of Table I reveals that several virus coat proteins and leucine zipper motif containing proteins enter prominently into this list. For the latter proteins, it has been shown that binding and folding are intimately coupled, i.e., that many proteins in the leucine zipper family lack an internal structure when being in the monomeric form.^{16,18,20} The structural analysis of chains listed in Table I indicates that predominantly either these form

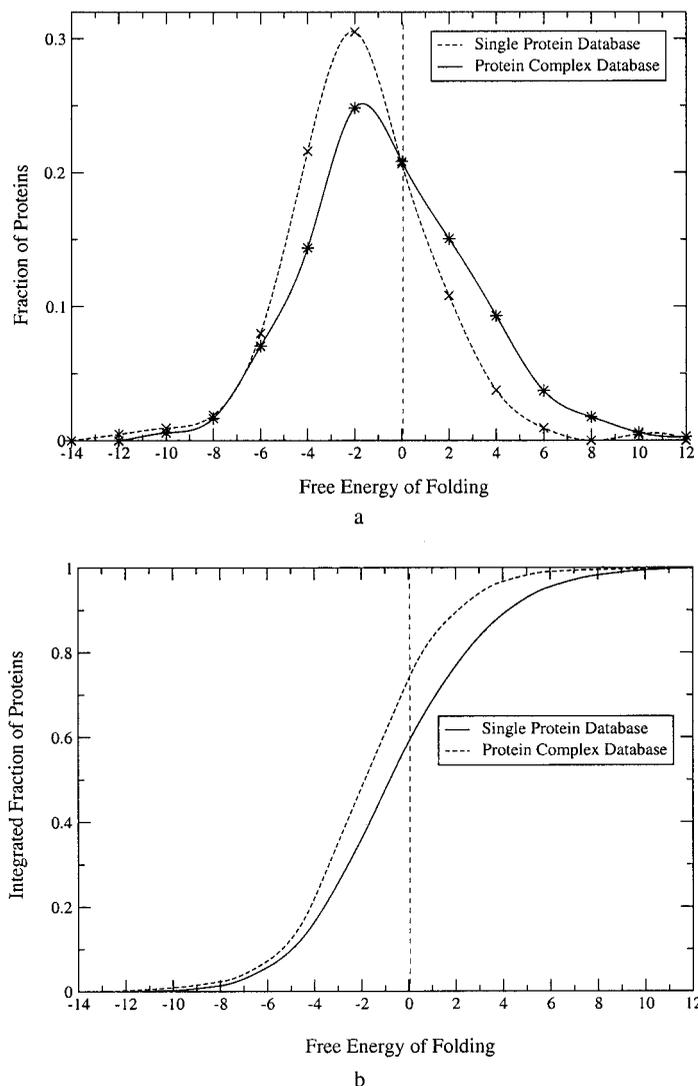


FIGURE 9 (a) A histogram of free energies of folding per residue at $T = 300$ K for 1024 individual proteins in the Protein Complex Database and 213 proteins in the Single Protein Database. (b) An integration of the corresponding curves in the upper graph.

long helical chains heavily interacting with other proteins, or they form nonglobular conformations which “make sense” only within the context of additional interactions with partner proteins.

An interesting deviation from the abovementioned trend is the crystal structure of the NH_2 -terminal domain of the lymphocyte cell adhesion protein, CD2 dimer (Figure 10). It turns out that the CD2 dimer is a metastable state that is not recovered when the protein is unfolded and then refolded.⁵⁰ Instead, a more stable monomeric structure emerges that has a very different fold than each monomer in the CD2 dimer. Although two domains in the CD2 dimer (Figure 10) resemble the single monomer domain, these domains are formed not by individual monomer do-

main but by intercalation of two protein chains (i.e., “fold swapping” instead of domain swapping).⁵⁰ Therefore, it is highly likely that individual monomer conformations as present in the CD2 dimer (Figure 10) are not viable native states when the other partner chain is not present, rationalizing the computed large positive free energy of folding.

From the previous analysis we have concluded that at least on the order of 15 per cent of monomeric chains in the Protein Complex Database are unfolded at room temperature while on their own. For a typical protein complex containing such a chain we have computed a representative phase diagram. We have carried out a similar calculation for the average database parameters as well (see Figure 11 and Table II).

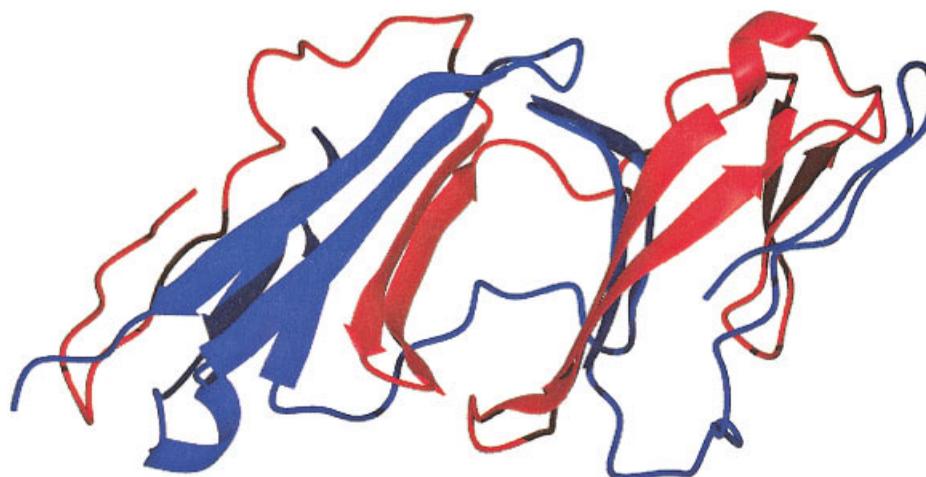
Table I Twenty Protein Chains from the Protein Complex Database With the Most Unfavorable Individual Free Energies of Folding (Membrane Proteins 1occ and 1prc Were Excluded from the List)

Protein	Chain	$\Delta F_{\text{folding}}$	Short Description
2bpa	3	7.23	Bacteriophage phix174 coat proteins
1una	A	7.32	Unassembled virus coat protein dimer
1lta	C	7.40	Heat-labile enterotoxin (lt) complex with galactose
1mec	4	7.48	Cardio picornavirus coat protein
1cdc	A	7.52	Cd2, N-terminal domain (1–99), truncated form
1tqx	B	7.60	Toxin gamma (cardiotoxin)
1mhl	A	7.75	Human myeloperoxidase isoform c
1cdc	B	7.88	Cd2, N-terminal domain (1–99), truncated form
1bbt	4	7.94	Foot-and-mouth disease virus
1mhl	B	8.25	Human myeloperoxidase isoform c
1tvx	B	8.45	Neutrophil activating peptide-2 variant form m6l
1tqx	A	8.87	Toxin gamma (cardiotoxin)
1fos	G	9.24	Two human c-fos : c-jun : dna complexes
2zta	B	9.67	Leucine zipper monomer
2zta	A	10.26	Leucine zipper monomer
1got	G	10.77	gt-alpha/gi-alpha chimera and the gt-beta-gamma subunits
1lya	A	11.04	Lysosomal aspartic protease, cathepsin d
1fle	I	11.42	Elafin complexed with porcine pancreatic elastase
1tmf	4	12.44	Theiler's murine encephalomyelitis virus coat protein
1lpb	A	13.32	Lipase complexed with colipase

The folding energy gap in case B is destabilized compared with the average database gap, but still is well within one standard deviation from the average ($-11.8 \text{ kJ} \pm 2.9 \text{ kJ/residue}$). The binding gap, on the other hand, was set to be stabilized by one standard deviation from the database average (Table II).

The analysis of the reference phase diagram for the average database parameters (see Figure 11a) reveals that the protein largely folds first before proceeding with binding. When looking at the $Q_b = 0$ projection

in Figure 11a it may seem at first surprising that the minimum in free energy function occurs at $Q_f = 0.6$ instead of $Q_f = 1$ as for the properly folded chain. Notice, however, that $F(Q_f = 1)$ is lower in free energy than $F(Q_f = 0)$; the shape of the curve for the intermediate Q values is dictated by a delicate balance of entropy and variance on one hand and the native energy gap on the hand. The reason that we obtain a minimum in free energy for intermediate Q instead of a maximum (i.e., folding barrier) is probably due to

**FIGURE 10** The representation of the CD2 dimer crystal structure. Monomers A and B are drawn in red and blue, respectively.

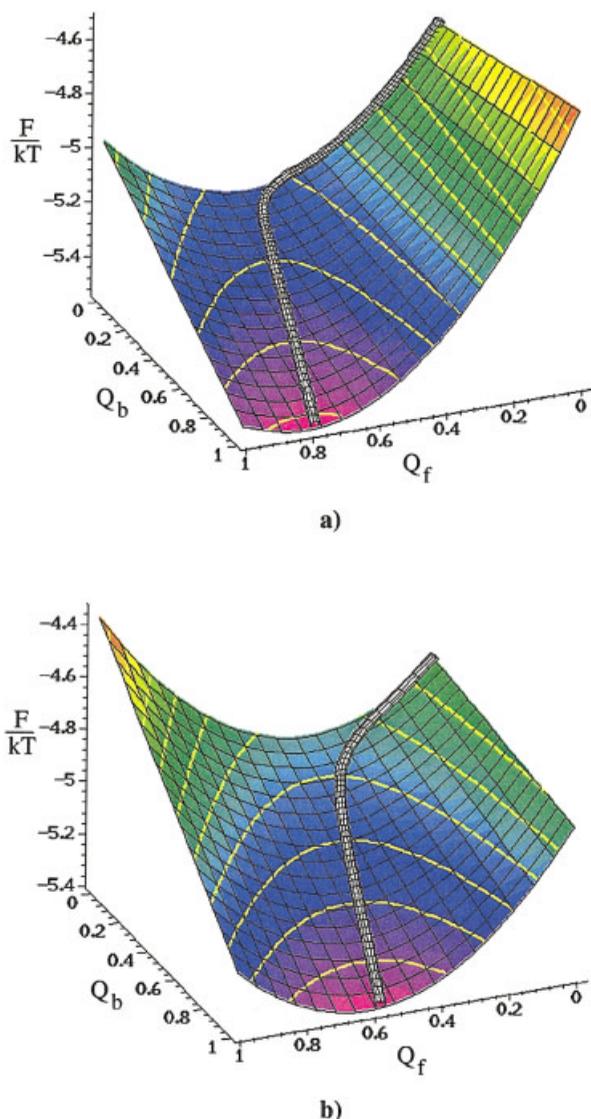


FIGURE 11 Free energy as a function of folding and binding order parameters. (A) Average database parameters. (B) Destabilized folding and stabilized binding parameters.

the quasilinear approximation for entropy which we have adopted in this work. Since we explore currently only thermodynamic aspects of various phase coexistence in the Protein Complex Database, the absence of barriers does not play an important role in our analysis. If our theory is to be used to investigate the

kinetic aspects of coupling of binding and folding, then more refined entropy model needs to be devised, most likely based on polymer physics arguments (a polymer chain entropy model by Plotkin and Wolynes may serve as a starting point³¹). We also note that nonadditive contributions to the free energy such as hydrophobic force and sidechain packing may also give rise to barriers.

As for the proteins that are more flexible while on their own, the binding and folding coordinates get coupled more intimately (Figure 11b). Notice that $F(Q_f = 1, Q_b = 0)$ vertex is higher in free energy than $F(Q_f = 0, Q_b = 0)$, i.e. the protein is expected to be unfolded while isolated. The native interface vertex, $F(Q_f = 0, Q_b = 1)$, is found to be slightly lower in free energy than the completely unfolded state $F(Q_f = 0, Q_b = 0)$. This does not rule out a possibility of interface contact ordering as a first step in the binding-folding mechanism. A study of such a possibility in detail is best left to capillary style analytical theory, the subject of work in progress.⁵¹ Finally, $F(Q_f = 1, Q_b = 1)$ is found to be significantly lower in free energy than the unfolded state, suggesting that an unfolded protein folds when simultaneously native binding contacts start forming.

Certain conditions are to be met for this happen, however. For instance, the relative ratio of binding to folding contacts (i.e., roughly speaking the interface surface area to protein volume ratio) must not be too low. In addition, the binding funnel must be deeper than the average. Within our simple theory of coupling binding and folding, one may use Eq. (10) in an appropriately parameterized form to explore the likelihood of particular geometrical and energetic parameters to induce folding upon binding.

CONCLUSIONS

As ever increasing amount of evidence is being presented for the considerable role played by highly flexible proteins and peptides in biological processes, signaling, and transcription being among prime examples, rationalizing the functional significance of such flexibility has become an important objective in current biological research. Inspired by the success of the

Table II Various Parameters Used to Compute Phase Diagrams A and B in Figure 11

	T	β	γ_f	γ_b	S^0	$S(0, 1)/S(0, 0)$	$\text{Gap}_{\text{Folding}}$	$\text{Gap}_{\text{Binding}}$	$\sqrt{\text{Var}_{\text{Folding}}}$	$\sqrt{\text{Var}_{\text{Binding}}}$
A	300 K	0.2	1.0	0.0	$3 k_B$	0.75	-11.8 kJ	-9.3 kJ	4.3 kJ	3.4 kJ
B	300 K	0.2	1.0	0.0	$3 k_B$	0.75	-10.3 kJ	-12.9 kJ	4.3 kJ	3.4 kJ

energy landscape viewpoint to rationalize the mechanism of protein folding, we have extended in this work that concept to construct a simple quantitative theory of coupling of folding and binding.

Specifically, the stratum Generalized Random Energy Model as applied to protein folding by Plotkin and Wolynes has served as a starting point for building our model. In our model, an unfolded protein transiently associates with a rigid surface of a folded partner protein, possibly followed by ordering of native interface contacts and folding of the protein chain. We have derived a free energy expression as a function of folding and binding order parameters Q_f and Q_b [Eq. (10)], which are useful in monitoring the progress of possible phase transitions. A nonadditive entropy term in the free energy expression tightly couples binding and folding coordinates. We have employed a quadratic quasilinear interpolation for entropy as a function of binding and folding order parameters, by estimating entropy at four vertexes of the Q_f and Q_b domain. Polymer physics ideas have been used to calculate the reduction in entropy when the native binding contacts are pinned at the partner protein's surface but the rest of the protein is still unstructured.

The binding and folding energy gaps as well as the corresponding energy variances entering into the free energy expression have been determined by statistical analysis of a large protein complex database compiled by Ben-Tal and co-workers. Miyazawa–Jernigan contact potential has been used to calculate the energies for individual protein conformations. The average energy and the energy variance of the denaturated state has been modeled by the sequence permutation of the database proteins. While the protein folding parameters have been derived without much surprise, the binding energy gaps were found to be strongly modulated by interface composition. In particular, hydrophobic interfaces were determined to be funneled to nearly the same degree as in the protein folding case, yet hydrophilic interfaces were found to be antifunneled. We have speculated that water-mediated interactions may play a role; the details on the corresponding potential will be published elsewhere.

Since the coupling of binding with folding makes much sense only when one of the proteins is (partially) unfolded while on its own, it is desirable to estimate how common these unstructured proteins are in the database. Consequently, we have computed the fraction of unfolded proteins being destabilized by at least 1 kJ/residue, which turn out to be 31% in the Protein Complex Database. Comparing this number with 16% calculated for a Single Protein Database we have suggested that at least on the order of 15% of protein

chains in the former database are unfolded (or adopt very different native conformations) in the absence of partner protein interactions. Using the parameters characteristic of these unstructured proteins, we have calculated free energy as a function of binding and folding order parameters. The analysis of the latter phase diagram has suggested that indeed an unfolded protein may fold when the formation of sufficiently stabilizing native binding contacts, which are *consistent* with the folded conformation, is initiated. We hope that the basic quantitative theory of coupling of folding and binding developed in the current work will serve as a useful stepping stone as well as a reference point for more elaborate models.

We gratefully acknowledge the National Institute of Health for its generous support of this work through a National Institute of Health Postdoctoral Fellowship Award to Gargegin A. Papoian. The effort of Peter G. Wolynes in concepts of protein folding is supported through NIH grant 5R01GM44557. We thank M. P. Eastwood, V. Lubchenko, S. S. Plotkin, and J. Ulander for very stimulating discussions. We are also pleased to contribute this paper to the celebratory volume honoring Schneior Lifson, a pioneer in bringing quantitative theoretical ideas to bear on molecular biology.

REFERENCES

1. Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing, New York: 1999.
2. Fischer, E. *Ber Dtsch Chem Ges* 1890, 23, 2611.
3. Fischer, E. *Ber Dtsch Chem Ges* 1894, 27, 2985.
4. Koshland, D. E., Jr. *Angew Chem Int Ed Engl* 1994, 33, 2375–2378.
5. Wright, P. E.; Dyson, H. J. *J Mol Biol* 1999, 293, 321–331.
6. Dyson, H. J.; Wright, P. E. *Curr Opin Struct Biol* 2002, 12, 54–60.
7. Hart, T. N.; Read, R. J. *Multiple-Start Monte Carlo Docking of Flexible Ligands*; Birkhauser: Boston, 1994.
8. Schaffer, L.; Verkhivker, G. M. *Protein Struct Funct Genet* 1998, 33, 295–310.
9. Verkhivker, G. M.; Bouzida, D.; Hehlhaar, D. K.; Rejto, P. A.; Arthurs, S.; Colson, A. B.; Freer, S. T.; Larson, V.; Luty, B. A.; Marrone, T.; Rose, P. W. *J Comput-Aid Mol Des* 2000, 14, 731–751.
10. Northrup, S. H.; Erickson, H. P. *Proc Natl Acad Sci USA* 1992, 89, 3338–3342.
11. Janin, J. *Proteins Struct Funct Genet* 1997, 28, 153–161.
12. Zhang, C.; Chen, J.; DeLisi, C. *Proteins Struct Funct Genet* 1999, 34, 255–267.
13. Fraunfelder, H.; Sliger, S. G.; Wolynes, P. G. *Science* 1991, 254, 1598–1603.

14. Miller, D. W.; Dill, K. A. *Protein Sci* 1997, 6, 2166–2179.
15. Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci* 1999, 8, 1181–1190.
16. Sauer, R. T. *Nature* 1990, 347, 514–515.
17. Spolar, R. S.; Record, M. T., Jr. *Science* 1994, 263, 777–784.
18. Johnson, N. P.; Lindstrom, J.; Baase, W. A.; von Hippel, P. H. *Proc Natl Acad Sci USA* 1994, 91, 4840–4844.
19. Johnson, C. R.; Morin, P. E.; Arrowsmith, C. H.; Freire, E. *Biochemistry* 1995, 34, 5309–5316.
20. Zitzewitz, J. A.; Bilsel, O.; Luo, J.; Jones, B. E.; Matthews, C. R. *Biochemistry* 1995, 34, 12812–12819.
21. Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E. *Proc Natl Acad Sci USA* 1996, 93, 11504–11509.
22. Ozawa, T.; Nogami, S.; Sato, M.; Ohya, Y.; Y., U. *Anal Chem* 2000, 72, 5151–5157.
23. Uverski, V. N. *Protein Sci* 2002, 11, 739–756.
24. Romero, P.; Obradovic, Z.; Li, X.; Garner, E.; Brown, C. J.; Dunker, A. K. *Proteins Struct Funct Genet* 2001, 42, 38–48.
25. Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc Natl Acad Sci USA* 2000, 97, 8868–8873.
26. Terada, T. P.; Sasai, M.; Yomo, T. *Article* 2002, 99, 9202–9206.
27. Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. *Proc Natl Acad Sci USA* 1995, 92, 3626–3630.
28. Onuchic, J. N.; Luthey-Schulten, Z. A.; Wolynes, P. G. *Ann Rev Phys Chem* 1997, 48, 545–600.
29. Glaser, F.; Steinberg, D. M.; Vakser, I. A.; Ben-Tal, N. *Proteins Struct Funct Genet* 2001, 43, 89–102.
30. Plotkin, S. S.; Wang, J.; Wolynes, P. G. *J Chem Phys* 1997, 106, 2932–2948.
31. Plotkin, S. S.; Wang, J.; Wolynes, P. G. *Phys Rev E* 1996, 53, 6271–6295.
32. Eastwood, M. P.; Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. *IBM J Res Dev* 2001, 45, 475–497.
33. Eastwood, M. P.; Wolynes, P. G. *J Chem Phys* 2000, 114, 4702–4716.
34. D'Aquino, J. A.; Freire, E.; Amzel, L. M. *Proteins Struct Funct Genet Suppl* 2000, 4, 93–107.
35. D'Aquino, J. A.; Gomez, J.; Hilser, V. J.; Lee, K. H.; Amzel, L. M.; Freire, E. *Proteins Struct Funct Genet* 1996, 25, 143–156.
36. Lee, K. H.; Xie, D.; Freire, E.; Amzel, L. M. *Proteins Struct Funct Genet* 1994, 20, 68–84.
37. Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*; Oxford Science Publications: New York, 1986.
38. Portman, J. J.; Takada, S.; Wolynes, P. G. *J Chem Phys* 2001, 114, 5069–5081.
39. Carslaw, H. S.; Jaeger, J. C. *Conduction of Heat in Solids*; Oxford at the Clarendon Press: Oxford, 1959.
40. Miyazawa, S.; Jernigen, R. L. *J Mol Biol* 1996, 256, 623–644.
41. Zhang, L.; Skolnick, J. *Protein Sci* 1998, 7, 112–122.
42. Betancourt, M. R.; Thirumalai, D. *Protein Sci* 1999, 8, 361–369.
43. Vendruscolo, M.; Mirny, L. A.; Shakhnovich, E. I.; Domany, E. *Proteins Struct Funct Genet* 2000, 41, 192–201.
44. Bastolla, U.; Farwer, J.; Knapp, E. W.; Vendruscolo, M. *Proteins Struct Funct Genet* 2001, 44, 79–96.
45. Vendruscolo, M.; Najmanovich, R.; Domany, E. *Phys Rev Lett* 1999, 82, 656–659.
46. Vendruscolo, M.; Kussel, E.; Domany, E. *Fold Design* 1997, 2, 295–306.
47. Pacios, L. F. *J Chem Inf Comput Sci* 2001, 41, 1427–1435.
48. Papoian, G. A.; Ulander, J.; Wolynes, P. G. To be published.
49. Chang, I.; Cieplak, M.; Dima, R. I.; Maritan, A.; Banavar, J. R. *Proc Natl Acad Sci USA* 2001, 98, 14350–14355.
50. Murray, A. J.; Lewis, S. J.; Barclay, A. N.; Brady, R. L. *Proc Natl Acad Sci USA* 1995, 92, 7337–7341.
51. Papoian, G. A.; Plotkin, S. S.; Wolynes, P. G. Work in progress.