Variationally Determined Free Energy Profiles for Structural Models of Proteins: Characteristic Temperatures for Folding and Trapping †

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2008

Impact Factor: 3.3 · DOI: 10.1021/jp076280n · Source: PubMed

4 AUTHORS, INCLUDING:

Peter Wolynes
Rice University
359 PUBLICATIONS **29,010** CITATIONS

Variationally Determined Free Energy Profiles for Structural Models of Proteins: Characteristic Temperatures for Folding and Trapping†

Tongye Shen,‡ Chenghang Zong,§ John J. Portman,⊥ and Peter G. Wolynes*,‡

Department of Chemistry & Biochemistry, University of California at San Diego, and Center for Theoretical Biological Physics, La Jolla, California 92093-0371, Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, Department of Physics, Kent State University, Kent, Ohio 44240

Received: August 6, 2007; In Final Form: November 16, 2007

Characterizing the phase diagram for proteins is important both for laboratory studies and for the development of structure prediction algorithms. Using a variational scheme, we calculated the generic features of the protein thermostability over a large range of temperatures for a set of more than 50 different proteins using a model based on native structure alone. Focusing on a specific system, protein G, we further examined, using a more realistic model that includes the nonnative interaction, the thermostability of both the native state and a collection of trap structures. By surveying the native structures for many proteins and by paying closer attention to the various trap structures of protein G, we obtained an overall understanding of the folding dynamics far from the conditions usually focused on; namely, those near the folding temperature alone. Two characteristic temperatures (shown to scale with folding temperature in general) signal drastic changes in the folding mechanism. The variational calculations suggest that most proteins would, indeed, fold in a barrierless manner below a critical temperature analogous to a spinodal in crystallization. For fixed interaction strengths, this temperature, however, seems to be generally very low, ∼50% of the equilibrium folding temperature. Likewise, native proteins, in general, would unfold in a completely barrierless way at a temperature 25% above folding temperature according to these variational calculations. We also studied the distribution of free energy profiles for escape from a set of trap structures generated by simulations.

I. Introduction

How protein molecules form three-dimensional and functioning structures has been the focus of many years of experimental and theoretical biophysics research.[1−4] Proteins are a very special class of heteropolymers. First, proteins are at the small size end of the large family of polymers which should have been well characterized by general polymer physics principles.[5−7] Second, and more important, evolutionary pressures have sculpted the unique heterogeneity intrinsic in the natural sequences to make each specific protein structurally different from others, allowing specificity in biochemistry. These features make protein physics quite different from the general polymer physics appropriate to most artificial polymers. Nevertheless, the dominant folded structures have been demonstrated a long time ago to be directly the physical consequence of the sequence information stored within the molecule and thereafter result from physical interactions defined by the sequence chemistry.

Studying the phase diagrams of proteins provides an organized way of viewing their folding mechanisms. It is especially important for several hotly debated topics: "downhill" folding,[8,9] the existence of "natively unfolded" proteins,[10,11] and perhaps the origin of misfolding-induced diseases.[12,13] Viewing protein molecules as physical objects at the coarse-grained level with a position resolution of each monomer, one can select a set of parameters that describe how close a particular configuration of protein is to the native structure in each of its regions. This allows a folding free energy profile to be drawn as a function of such folding order parameters. Once the free energy profiles under various environmental parameters are given, kinetic properties can also be computed.[14−16]

The majority of the previous folding studies, especially those using simulation, have focused on the behavior around the equilibrium folding temperature. Only a few idealized models with a generic protein structure have ventured into a wider range of temperatures, such as those studied in refs 17−20. Here, we inspect features of a free energy landscape under a wide range of stabilitites of folded and realistic misfolded structures. Typically, that temperature, T_f, is defined for a two-state folder as the temperature at which the unfolded and the folded ensembles have the same free energy. The corresponding free energy profile is characterized by a barrier separating two basins of equal depth. When the temperature increases, the unfolded basin is lowered relative to the native basin, and vice versa as temperature decreases. Sometimes in vivo or in vitro, a larger range of temperature, denaturant concentration, or other perturbation may alter more dramatically the stability of these functioning structures. If the stabilities of the two basins can be sufficiently changed, one can imagine, in analogy to spinodal phenomena seen in material science, that at some point, the metastable basin will be sufficiently high on the profile that the metastable basin will completely disappear. Consider the sketch of the two-state profile shown in Figure 1. Here, there is a critical downhill temperature T_d ($<T_f$) below which the free energy profile suggests folding becomes completely downhill to native basin. Similarly, there may be a critical activation temperature, T_h ($>T_f$), above which we have a profile that is

† Part of the "Attila Szabo Festschrift".
‡ University of California at San Diego and Center for Theoretical Biological Physics.
§ University of Illinois at Urbana-Champaign.
⊥ Kent State University.

10.1021/jp076280n CCC: $40.75
Published on Web 04/01/2008

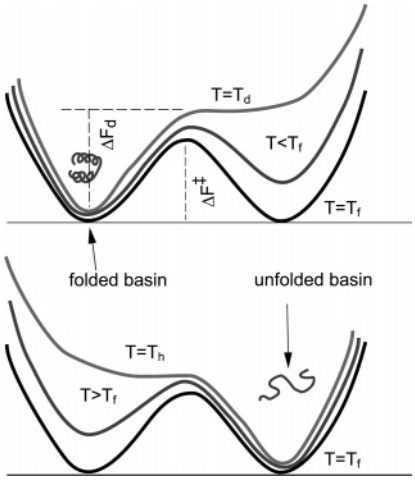

Figure 1. An illustration of the free energy profiles of a two-state folder at various characteristic temperatures.

completely downhill to the unfolded basin due to the dominance of entropy. The approach to this behavior is experimentally evident in the movement of the transition state, which typically upon changing temperature moves away from the stable state, in accord with Hammond behavior until it would be extrapolated to merge the metastable minimum at T_d or T_h.

Apart from making transitions between the native basin and an undifferentiated set of extended structures, protein molecules may sometimes explore long-lived compact structures other than the native. Such compact structures having a very different set of contacts from the native are called the misfolded "trap" structures. Such behavior is the hallmark of glassy dynamics. Previous analogues suggest that such trapping is rarely seen near T_f for evolved proteins.[21−23] This is the quantitative statement of the minimal frustration principle.

Nevertheless, misfolded structures are thought by some to cause a class of diseases.[12] A trap structure with high stability may potentially present two problems for living cells: (1) Such misfolded structures may be specifically harmful in some way, interfering with the normal biological functions of other components of the cell. For example, they may have the structure of ion channels puncturing the membrane. (2) They may kinetically prevent the protein from folding to the correct structures before being cleared by house-keeping mechanism; thus, they may interfere with the normal function of that protein. Even if traps do not cause disease by being accessible under real physical conditions (with the physically correct, natural energy function), they do present some difficulties for structure prediction algorithms in which the energy function is only approximately known.[24,25] Thus, it is important to study the free energy profiles for forming these trap structures. In this paper, we exploit the fact that the same statistical mechanical concepts characterizing the structural stability of the native state can be applied to the investigation of traps, as well.

Obviously, it is important to include nonnative interactions when we want to study glassy dynamics and the stability of misfolded structures. There have been a few studies exploring quantitatively the effects of nonnative interactions. Instead, the majority of the folding free energy profiles calculated for specific protein structures have been based on a Hamiltonian having native-only interactions. The validity of such calculations is based on the minimal frustration principle,[22] which suggests that interactions between residues that do not form contacts in the native structure are less important than those that do. By virtue of the minimal frustration principle, native-only models provide the zero-order solution of the folding problem, whereas the effect of nonnative interactions is to act as a friction for motion on the perfect funnel.

In this paper, we will use a variational free energy functional method[14,15] to sketch the phase behavior[17,26] for forming both native structures and traps. Variational methods are convenient because they evaluate both the energy and the entropy components of stability on an even footing. Being a numerical partition function based approach, the variational scheme does not have statistical errors of simulation. We detail the Hamiltonians used here and the specific variational scheme in the next section (Section II). We then report the characteristics of the free energy profile found from the native-only Hamiltonian for a survey of more than 50 proteins (listed in the format of PDB_code-chain_index:[27] 1BBL, 1BNI-A, 1BRS-D, 1BX7, 1C9O-A, 1CSP, 1EBD-C, 1F94-A, 1G8Q-A, 1GVP, 1HZ6-A, 1JY2-O, 1KAF-A, 1KQ1-A, 1LMB, 1N0Q-A, 1N13-A, 1OAI-A, 1PGB, 1PRB, 1R0R-I, 1R69, 1RIS, 1SHG, 1TEN, 1TIT, 1TUK-A, 1UBQ, 1URN-B, 1UTG, 1VII, 1VFY-A, 1WHZ-A, 1W4H-A1, 1YBK-A, 1ZV1-A, 1ZVA-A, 256B-A, 2B5A-A, 2BAY-A, 2BWF-A, 2CYU-A1, 2CI2-I, 2CRO, 2CS7-A, 2D68-A, 2DS5-A, 2G6F-X, 3CHY, and 4PTI) in Section III. Afterward, we discuss the effects of nonnative interactions on both the folding and trapping free energy profiles studying a specific protein, protein G (PDB code 1PGB,[28] a small alpha/beta protein of 56 residues). Here, we also use the variational method to calculate the stability of various misfolded (traps) or partially folded (on pathway intermediates) structures in addition to the native structure of this protein in Section IV. Finally, we conclude in Section V.

II. Methods

Previous studies using the variational scheme[14,15] have shown good agreement between theory and experiment for the overall folding rates for a set of proteins[29] and have provided very good, detailed comparisons of the structures formed along free energy profiles, as judged by studying the kinetics of multiple mutants of specific proteins.[30,31] These studies confirm that the simple model Hamiltonian and the scheme of approximation we adopt are, at least in the part of the phase diagram near equilibrium folding, adequate. The variational scheme treats the folding and unfolding processes as structural localization and delocalization transitions, such as freezing and melting of a solid. A similar variational approach for computing the stability and dynamics of collections of self-organized particles has been previously successfully applied to study excluded-volume-driven solid phases,[32] tensional-driven elastic phases,[33] and even nonequilibrium motorized structures.[34]

In the variational method, a protein is modeled as a collection of beads, each representing one monomer of the polymer, that is, the amino acid residue. A backbone energy term, H_B, prevents the breaking of the polymer chain, and the more interesting part of the Hamiltonian, H_I, describes the interaction (contact energy) between monomers. The present calculations use only a pairwise summation over interactions between different amino acid residues (beads).

The backbone Hamiltonian, H_B, has a simple quadratic form that gives correct persistence lengths:

$$H_B = \sum_{i=1}^{N-1} D_1[\vec{r}_i - \vec{r}_{i+1}] + \sum_{i=1}^{N-2} D_2[\vec{r}_i, \vec{r}_{i+1}, \vec{r}_{i+2}]$$

Here, the chain length, N, is the total number of the amino residues of the protein chain. The functions D_j are quadratic

functions of residue positions. By adjusting the coefficients to give the observed persistence lengths for protein, this quadratic Hamiltonian mimics (semiflexible) polymers fixing the bond-angle level.[35] Following the various previous applications of the scheme, the backbone energy term itself is treated as entropic and therefore thermal; that is, $H_B(T) \propto k_B T$. Thus, in our calculation, βH_B is kept fixed as temperature is varied. The final probability distribution of the noninteracting polymer (when $H_I = 0$) is independent of temperature.

The total contact energy is a summation over pairs, that is, $H_I = \sum_{ij} u_{ij}$. Here, the interaction u_{ij} is between those residues i and j that are not sequence neighbors, that is, $|i - j| \geq 3$. Unless the u_{ij} were taken to be a quadratic of positions of residues i and j, generally, the partition function for a system having the total Hamiltonian $H = H_B + H_I$ cannot be solved exactly. Instead, using the variational approximation, we assume that a set of the trial functions for the polymer configuration distribution can be characterized by restrained ensembles of the form $p^v(r) \propto \exp[-\beta H^v(\{\vec{r}_1, \vec{r}_2, ..., \vec{r}_N\})]$. Such a trial Hamiltonian is taken to be a summation of single-body terms that localizes each amino acid around its native position (or localizes residues to positions in a specific trap in the case in which we study the properties of a trap's escape free energy profile), $\vec{r}^*$:

$$H^v = H_B + \sum_j 3k_B T/(2a^2) C_j (\vec{r}_j - \vec{r}_j^*)^2$$

Since every term here is up to quadratic of position, we can reexpress more explicitly this feature as

$$\beta H^v = \sum_{ij} 3/(2a^2)[G^{-1}]_{ij} (\vec{r}_i - \langle \vec{r}_i \rangle)(\vec{r}_j - \langle \vec{r}_j \rangle) + \text{const.}$$

Here, matrix G is the covariance matrix of the thermal fluctuation at the residue level. We set the length scale parameter to $a = 3.8$ Å, and it makes G dimensionless.

The variational free energy resembles the standard thermodynamic perturbation theory. It is used to obtain the final free energy profile as a function of trial localization parameters, C_j; that is,

$$F[C] \simeq F^v[C] + \langle H - H^v \rangle_v$$

Here, $\langle \cdot \rangle_v$ is the average over the ensemble of the distribution $p^v(r)$. The above procedure transforms the initially difficult statistical mechanical problem of particles moving in 3D space to a free energy function of scalars for each residue, which can be evaluated exactly numerically. The original $3 \times N$ degrees of freedom are the Cartesian coordinates representing individual *microscopic* configurations, whereas the trial functions having N constraint values each specify an *ensemble* of structures; that is, $H(\{\vec{r}\}) \rightarrow F(\{C\})$. In $F[C]$, the parameters C_i (total number is N) have the meaning of Debye−Waller factors. Each C_i indicates how localized a given residue i is around the native state position $\vec{r}_i^*$ in the restricted trial ensemble of configurations. As mentioned above, this type of order parameter has also been used to describe a simple localization transition of other soft matter systems, such as the cytoskeleton. The ideas behind the variational method are sketched in the Figure 2.

We apply the variational scheme to two different contact Hamiltonians, H_I. First, we study a native-contact-only approximation, $H_I = H_{NC}$. In this Hamiltonian, only native interactions are considered; that is, one sets $u_{ij} \equiv 0$ if i and j do not form a contact in the native state. Otherwise, the form of the function is $u_{ij} = u(r_{ij})$; that is, it depends on the distance

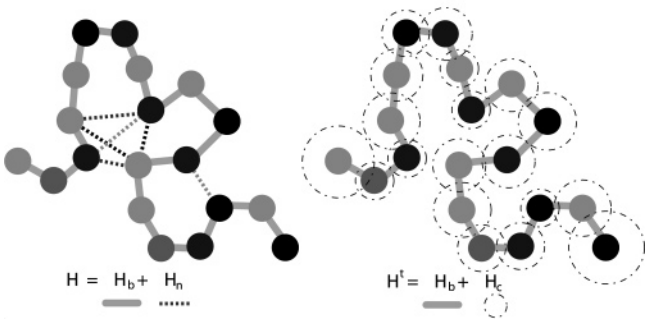

Figure 2. A sketch of the variational method for the Hamiltonian of protein systems. We replace the target Hamiltonian with a solvable Hamiltonian expressed as a quadratic function of the positions of the monomers. The coefficients of the quadratic function determine how localized each residue is around its mean position.

between the two residues. Here, a native contact between residue i and j is regarded as being formed if there is any side chain or backbone atoms (except hydrogens) of the two amino acid residues based on the native structure that are within 4.0 Å of each other.

The actual form of the pair potential, $u_{NC}(r) = \sum_k \gamma_k \exp(-\alpha_k r^2)$, is a sum of three Gaussian terms constructed to mimic a contact potential that is repulsive at short distance, attractive at intermediate distance, and eventually levels off. The detailed form of the potential with its parameters was reported by Portman et al.[15] The energies of the specific native contact pairs were obtained from the Miyazawa−Jernigan potential.[36]

We also study, especially for investigating traps, another contact Hamiltonian, $H_I = H_{AC}$, in which all the nonneighboring pairs are allowed to have interactions given by a transferable function of the chemical identity of the pair of the residues and the distance of the separation. The functional form of $u_{AC}(r) = \sum_k \epsilon_k \Theta(r - l_k)$ in these cases is simply a sum over three step-function terms. For the general pair interaction u^{ij} between residues i and j, we use the form of the interactions that has been used for predicting α and α/β proteins de novo.[37,38] These interactions were obtained via an optimization of a structure prediction Hamiltonian based on associative memory Hamiltonian approach.[39] The pair potential specifically is written as

$$u_{AC}^{ij}(|\vec{r}_{ij}|; \epsilon_{1,2}^{ij}, l_{1,2,3,4}) = \epsilon_0 + \Theta(r^{ij} - l_1) \times (\epsilon_1^{ij} - \epsilon_0) + \Theta(r^{ij} - l_2) \times (\epsilon_2^{ij} - \epsilon_1^{ij}) - \Theta(r^{ij} - l_3) \times \epsilon_2^{ij}$$

Here, energy $\epsilon_0 = 1$ kcal/mol. All the other parameters, $\epsilon_{1,2}$ and $l_{1,2,3}$, come from ref 39. Note here the coefficient $\epsilon_{1,2}$ depends on the chemical identity of the pair but all other terms are generic. The term $\epsilon_0 \times (1 - \Theta(r - l_1))$ is unique to the current model. It effectively establishes the steric effect, which had been realized by a nonspecific van der Waals interaction in the AMH simulation. Here, we have shortened the notation by setting $|\vec{r}^{ij}| = r^{ij}$.

Below, we tune through a series of Hamiltonian interpolating from a native-only to a Hamiltonian that includes nonnative interactions. We study a series of interaction models,

$$H(m) = H_B + H_{NC} + m \times H_{AC}$$

with $0 \leq m \leq 1$.

Before we go into the details of the calculations, we want to reemphasize what input information is needed to construct the various terms of the Hamiltonian for a *specific* given protein. The term H_B requires only quite generic parameters of polymer

physics that are not sensitive to protein details. The term H_{NC} requires the specification of the native contact information of the given protein; that is, one must supply a list of contact pairs of beads made in the native structure. The term H_{AC}, having transferable but heterogeneous parameters, depends on the sequence of the specific protein. The variational term $\sum_j C_j(\vec{r}_j - \vec{r}_j^*)^2$ requires the native position of the residues as the location of the fiducial sites for the trial function. When calculating the trap's escape free energy profile, the fiducial sites are supplied by the trap structure.

The details of evaluating $\langle H_{NC}\rangle$ from a sum of Gaussian forms can be found in Portman et al.[15] We only mention here the additional features involved in the calculation of $\langle H_{AC}\rangle$. In general, the average of a pair potential for a given site localization trial function is obtained from a pair distribution function, $\langle u\rangle^{ij} = \int d\vec{r}^{ij}\, u(\vec{r}^{ij})\, \rho(\vec{r}^{ij})$. The probability distribution for displacement of relative position $\vec{r}^{ij}$ from the mean, $\vec{R}^{ij}$, is

$$\rho(\vec{r}^{ij}) = (3/(2\pi a^2 \Delta G^{ij}))^{3/2} \exp(-(3(\vec{r}^{ij} - \vec{r}^{ij})^2)/(2a^2\Delta G^{ij}))$$

The relative displacement fluctuations provide $\Delta G_{ij} = G_{ii} + G_{jj} - G_{ij} - G_{ji}$, where G_{ij} are the elements of the inverse of the connectivity matrix defined in H^v. The "off-centered" Gaussian integrations involving $\Theta()$ can be written using the error function $\mathrm{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-t^2}\, dt$. The average takes the form

$$\langle u_{AC}\rangle = \epsilon_0 + \sum_{\alpha=1,2,3} (\epsilon_\alpha - \epsilon_{\alpha-1})\, f(l_\alpha) - \epsilon_3 f(l_4)$$

where the function f is given as

$$f(l) = 1 + \frac{1}{2}\mathrm{erf}\left[\frac{R-l}{\sqrt{\xi}}\right] - \frac{1}{2}\mathrm{erf}\left[\frac{R+l}{\sqrt{\xi}}\right] + \frac{(\xi/\pi)^{1/2}}{2R} e^{-(R-l)^2/\xi} - e^{-(R+l)^2/\xi}$$

We use the shorthand notation $\xi = 2a^2\Delta G/3$. Here $R = |\vec{R}^{ij}|$. We can obtain further analytical expressions for the first-order derivatives $\partial_R u$ and $\partial_\xi u$. From these equations, one can express the first-order derivatives $\partial\langle u_{AC}^{ij}\rangle/\partial C^k$ using the chain rule. The second-order derivatives can be also obtained similarly. We actually employ a numerical differentiation routine in the code implementation for the second-order calculation.

In order to study the stability of misfolded structures, we first generated a set of trap structures through coarse-grained molecular dynamics simulations. These traps are those characterized by the landscape of the associative memory Hamiltonian (AMH) previously employed for structure prediction.[40−42] The protocol for α/β protein structure prediction[37,38] was used in our study. In the case of true de novo structure prediction, multiple proteins, on whose native structures the target protein sequence has been threaded, are plugged in as "associative memories" to provide candidate local structural information. For the purpose of generating traps for the current study, we used the target native structure as the only memory in the AMH. This approximates the local (in sequence space) native signals. Furthermore, the transferable functions for the tertiary contact Hamiltonian, H_{AC}, were parametrized to agree with the long-range (again, sequencewise) contact potential in the AMC code. To generate traps during simulation, we quenched the temperature rapidly from configurations substantially above the folding temperature below which simulation usually falls into the native basin. We also used a bias potential in the quenching procedure to sample preferentially traps structurally far away from the native basin.

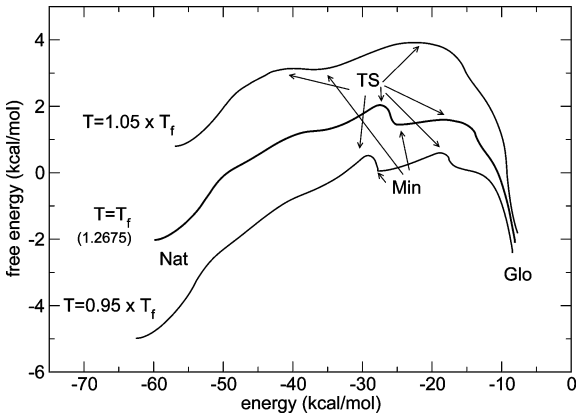

Figure 3. The folding free energy profiles near T_f for a native-only Hamiltonian of protein G.

III. Free Energy Profiles for Native-only Hamiltonians

We first present the free energy profiles for the native-only Hamiltonian $H(m = 0)$ around its folding temperature, T_f, for protein G. As shown in Figure 3, at the folding temperature, the unfolded and folded basins by definition have the same free energy. Mathematically, we were able to find two sets of free energy minima labeled by the transformed ensemble coordinates $\{C_i^N\}$ and $\{C_i^D\}$ (positioning the native and unfolded free energy minimum, respectively) satisfying $F[C^N] = F[C^D]$. These two deep minima are clearly distinguished by distinct folding order parameters. They are minima for different reasons: the $F[C^N]$ is a minimum due to the native contact energy formed between residues of protein, whereas the $F[C^D]$ is a minimum largely due to the contribution of entropy for an unfolded polymer ensemble. We use a global order parameter, defined as the total native contact energy, to display the free energies as a conventional one-dimensional plot, that is, $F[C]$ vs $E[C]$.

Throughout the paper, the energy units are taken to be kilocalories per mole and the temperature units are expressed in kilocalories per mole per the Boltzman constant. The folding temperature for protein G is $T_f = 1.268$. At T_f, we not only find two extremes corresponding to native and unfolded basins but also find two transition states (saddle points) and another shallow intermediate state basin that connect these two minima. The algorithm allows one to follow the two unstable directions of the saddle points $\{C_i^‡\}$ along the free energy gradient to the minima. The barrier height is $\Delta F^‡ = F[C^‡] - F[C^D] = F[C^‡] - F[C^N] = 4.11$. In Figure 3, we labeled the position of these critical points in the order parameter-free energy plot and linked them by recording the motions of the free energy downhill minimization trajectories. Note that the steps of the downhill motion are small enough for us to regard the resulting trajectory as describing an estimate of a quasiequilibrium ensemble of paths connecting the minima to the saddle points. This construction provides more information about the landscape than does the straight-line connection between critical points. We wish to stress again here that the "path" found by the algorithm is not a microscopic path of precise specific configurations, but rather, it describes a route of ensembles of structures. This ensemble path can be described by the transformed C representation.

Figure 3 also displays how small thermal perturbations change the landscape of protein G. When the temperature is raised by 5%, we see that the folded basin is raised relative to the unfolded

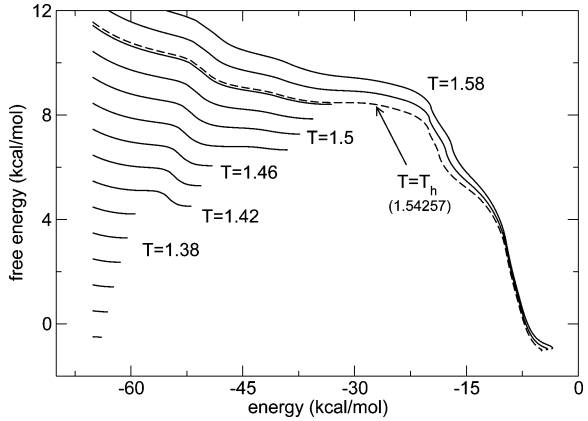

Figure 4. The onset of the uphill temperature T_h at $m = 0$ for protein G.

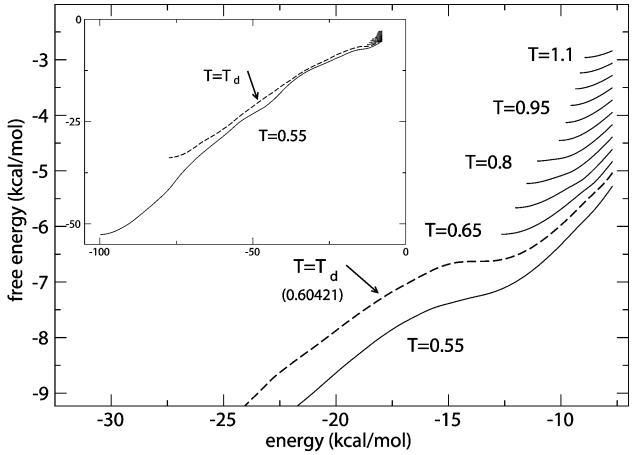

Figure 5. The onset of the downhill temperature T_d at $m = 0$ for protein G.

basin by an amount $F[C^N] - F[C^D] = 2.61 > 0$. While changing in the opposite direction, lowering the temperature by 5% makes the folded basin more stable by ~2.58 kcal/mol. The position of the native basin along the reaction coordinate, as indicated by the order parameter, $E[C^N]$, is more susceptible to thermodynamic perturbation than is the energy of the unfolded basin $E[C^D]$. This is quite reasonable because the unfolded basin is an entropy-stabilized minimum that contains little native energy content. Also from this result, we see the position of the major transition state along the reaction path shifts with temperature, moving toward the unfolded direction with higher temperature. This pattern was also previously observed in other proteins.[30]

When we started with the order parameters C^N for the native basin at T_f, gradually changed the temperature of the system, and tried to locate the new position of the native basin by free energy minimization, we found that the free energy of the folded basin rises and the basin location is found to move horizontally toward the unfolding direction. We find the new minimum C^* moves to a less folded state; that is, $E[C^*] > E[C^N]$, as shown in Figure 4. Eventually, at the critical temperature $T_h = 1.543$ for the case of protein G, the folded basin ceases to be even metastable. Above this temperature, the unfolded basin remains as the only minimum of the free energy profile. The movement of the native basin position is continuous with rising temperature until it reaches T_h. As shown in Figure 4, when tracing the position of the basin by scanning temperatures to cross T_h from below, the basin makes a discontinuous jump to the unfolded minimum along the native energy coordinate. We vary the temperature and record the critical temperature, T_h, that enables this abrupt change of landscape.

In a converse manner, when the temperature is lowered further below T_f, the denatured basin gradually shifts toward lower native energy, as shown in Figure 5. Again, the structural susceptibility of the denatured basin, $|\partial E_D/\partial T|$, is much smaller than that of the native basin at high temperature. Once the temperature is lowered to a spinodal temperature, T_d, which is ~0.618 for protein G, the profile becomes downhill from unstructured ensembles to the native basin. Below this spinodal temperature T_d, a system starting from the unfolded ensemble settles down at the minimum of the folded basin without needing thermal activation. In this model, the native basin has considerable entropy at T_f. Fewer conformational substates are availible as the temperature is lowered below T_f because the residues localize more tightly about the native conformation.

We carried out this analysis of the free energy profiles for a set of small proteins having chain lengths between 50 and 130. The set included more than 40 proteins of different folds and

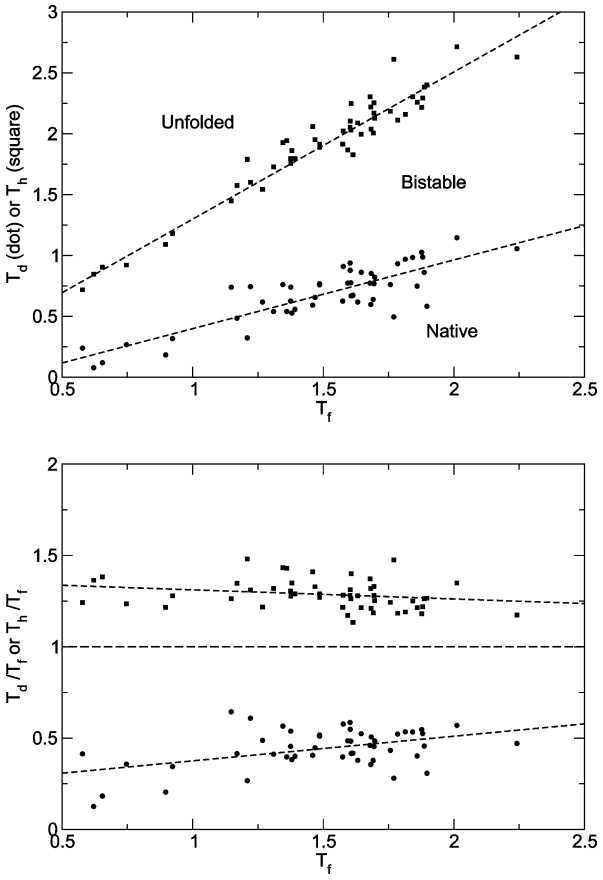

Figure 6. The characteristic temperatures for a set of more than 50 proteins.

several extra proteins whose kinetics have been well-characterized. After repeating the above procedure of locating the critical temperatures for the free energy profiles, we can plot the results to see a relation among the various characteristic temperatures: T_f, T_h, and T_d. We find these temperatures are correlated, as shown in Figure 6. Each symbol labels a specific protein's T_f vs its critical uphill or downhill temperatures, T_h or T_d. On average, we find $T_h/T_f \simeq 1.26$. There is a slight tendency for this ratio to decrease with protein size. On the other hand, for downhill spinodal barrierless folding, one finds $T_d/T_f \simeq 0.5$, with a slight tendency to increase with protein size. This almost universal nature of the ratio for models with a native-only interaction irrespective of the protein's fold was not completely

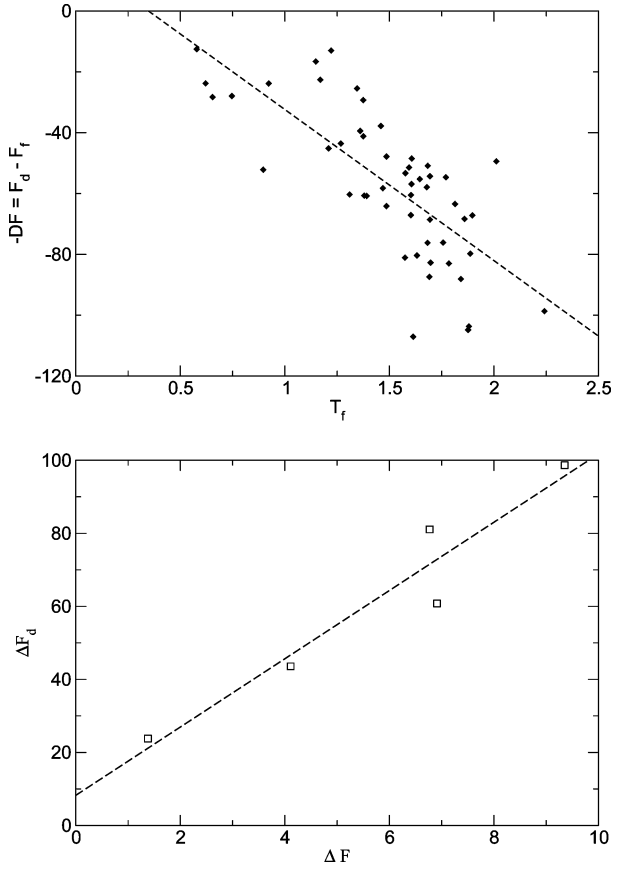

Figure 7. (a) The free energy drop at the onset of the downhill profile vs the folding temperature for a set of more than 50 proteins. (b) The comparison of the free energy drop at T_d versus the barrier height at T_f for five proteins listed in text.

expected by us. The relative variation of T_d/T_f from its mean is slightly larger than that of T_h/T_f.

The variational results give a range of two-state behavior. The range from T_d to T_f is larger than that from T_f to T_h. Making use of well-accepted approximate correspondences between theoretical calculations using thermal perturbations versus the experimental probe using chemical denaturant perturbation, the phase diagram can be translated to the experimental observables using the chevron plot: one expects to observe a stronger curved effect at the right wing (region of high denaturant concentration) as compared to the left wing, which according to the calculation persists to be linear over a larger range.

An interesting free energy scale that characterizes the downhill landscape, ΔF_d, has already been illustrated in Figure 1. More accurately, we define the drop of free energy for downhill landscape as the free energy difference between free energy of the folded basin at T_f and that found by the variational method method slightly below the downhill temperature, $T = T_d^-$, that is,

$$\Delta F_d = F_N(T_d^-) - F_N(T_f)$$

In Figure 7(a), we plot results of the drop of free energy, the needed extra stabilization for spinodal folding, for the same set of proteins. This free energy drop can be viewed as a measure of how difficult it would be to make a protein fully downhill by modifying the conditions via adding osmolytes etc. Note that ΔF_d turns out to be closely related to the kinetic barrier height, $\Delta F^{\ddagger}$, at T_f. Though it is quite difficult to calculate $\Delta F^{\ddagger}$ in an automatic fashion for the entire large protein set we used here,

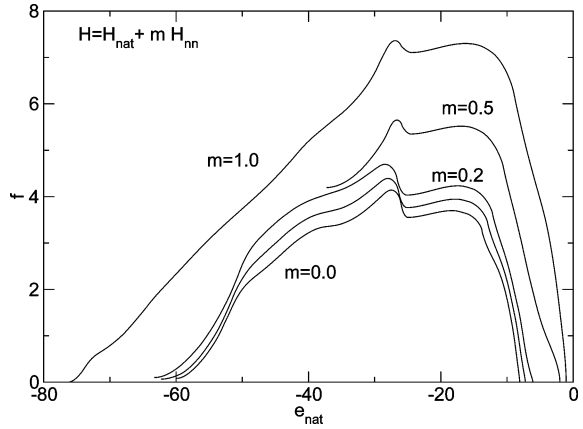

Figure 8. Free energy profiles of a series of Hamiltonians ($m = 0.0$, 0.1, 0.2, 0.5, and 1.0), including nonnative interactions of different strengths at the corresponding folding temperatures. The abscissa indicates the native energy component of the total energy.

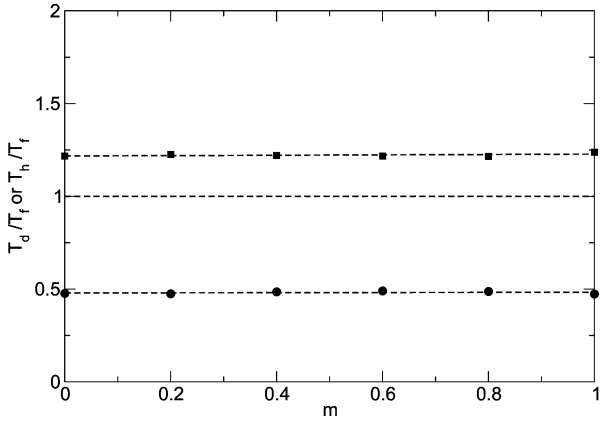

Figure 9. The ratio of the characteristic temperature (T_d or T_h) over the folding temperature as a function of the strength of the nonnative potential expressed by m.

for several proteins (2CYU, 1PGB, 1TIT, 1URN, and 2CRO) we have performed calculations of both $\Delta F^{\ddagger}$ and ΔF_d. These studies suggested a linear relation between these two quantities at the crudest level. More specifically, we found $\Delta F_d \simeq 10 \times \Delta F^{\ddagger}$ for the native-interaction-only Hamiltonian, as shown in Figure 7b.

Since the free energy barrier at T_f generally increases with chain length,[43] one can anticipate from this relationship that smaller proteins require a less biased environment to exhibit downhill folding or unfolding. According to the variational results, most large proteins should require rather extensive favorable environmental changes before their folding profiles would become downhill. The variational scheme suggests that only for very small proteins is achieving downhill folding by changing uniform external conditions a likelihood. Of course, this conclusion based on native-contact-only models assumes the protein is nearly completely devoid of frustration.

IV. Profiles with Nonnative Interactions for the Native Basin and Nonnative Traps

We briefly report on the effects of nonnative interactions on free energy profiles for both folding and trapping and for trap escape as predicted by the present variational scheme. To quantify the nonnative interactions, we use the tractable interpolating Hamiltonian with $m \neq 0$. We studied a range of Hamiltonians of the form $H(m) = H_B + H_{NC} + mH_{AC}$ with m

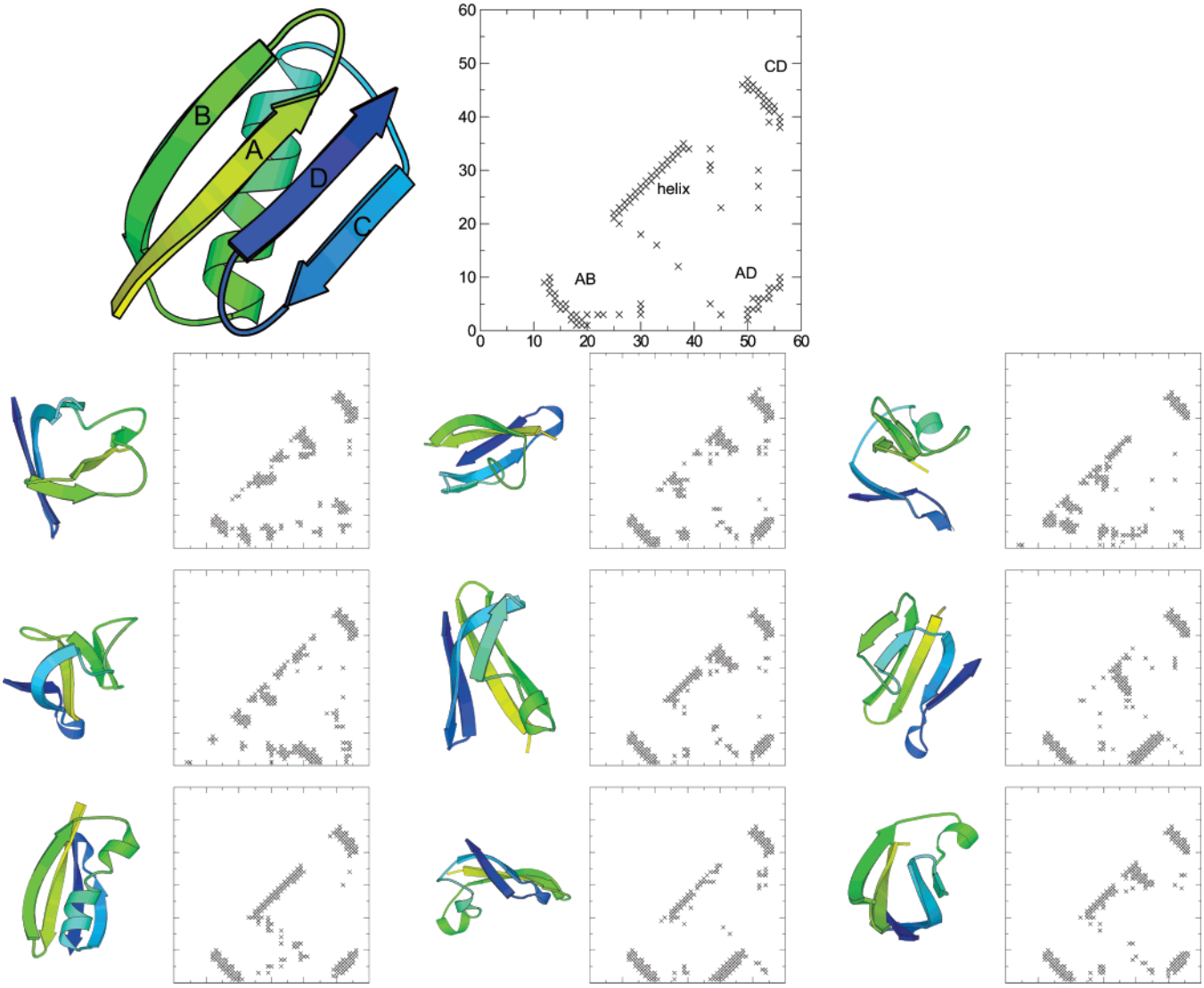

Figure 10. Protein G's native structure and nine examples of decoy (misfolded) structures.

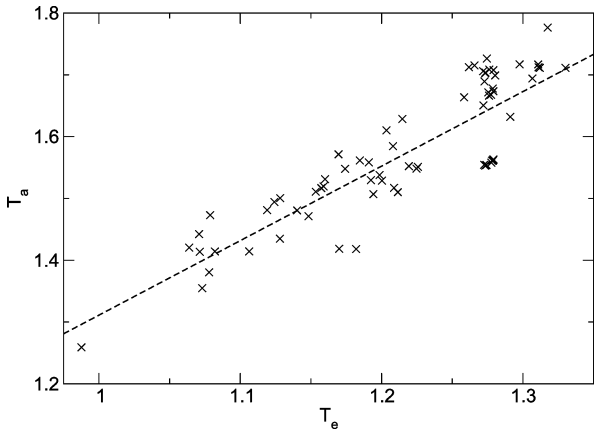

Figure 11. T_e versus T_a for a set of trap structures.

$= (0, 1]$. Again, here, H_B is the backbone chain interaction and H_{NC} is a native contact interaction, whereas H_{AC} is a transferable contact interaction summed over all the pairs with residue indices i and $j > i + 11$.

How do the nonnative interactions change the folding profile according to the variational approximation close to T_f? The results are shown in Figure 8. The overall thermal stability of the two basins around T_f does not change too much for the case of protein G. The greater energetic favorability of the native basin by H_{AC} is almost completely canceled out by a corresponding decrease in entropy. As a result, T_f hardly changes when we vary m; however, the barrier heights of the folding transition increase as nonnative terms are added. With increasing m, a higher barrier results, as shown in Figure 8. For a fully nonnative interaction with $m = 1$, the barrier height is 7.36. This is ~1.8 times that found with the native-only Hamiltonian. The free energy barrier height increases almost linearly with m.

There are several previous studies on the effect of nonnative interaction on the folding profiles with very idealized forms.[44−46] On the basis of the assumption that nonnative interaction is a very small amplitude, Gaussian, white noise, the conclusion of decreasing barrier height with increasing amplitude was drawn from these studies and is in direct contrast with our calculations showing barrier heights increased with added nonnative interaction. It is not difficult to see that our nonnative interaction, coming from a more realistic source (a generic physical interaction that is being used in actual structural prediction method) is neither of small amplitude (relative to native terms) nor of zero mean.[39] It is not surprising that the barrier height changes with extra interactions and depends on the specific forms of the interactions, and different assumptions of the nonnative interactions and whether total energy is normalized can lead to dramatically different results.

We again located T_d and T_h for $m \neq 0$. We plot the results in Figure 9. We find for protein G that the ratio T_d/T_f is around 0.5, that is, T_d is about 50% lower than T_f, regardless of m. On the other hand, T_h is ~25% higher than the corresponding T_f. From this, we again see one important point: that T_f is much closer to the "uphill" limit T_h than its distance to the temperature of downhill spinodal folding T_d. These two temperatures seem not to be very sensitive to m, even when the barrier height itself is sensitive to m.

For the full Hamiltonian, including nonnative interaction, we also calculated the drop of the free energy needed for it to reach T_d^-. It turns out that this value slightly increased to 47.4 kcal/mol from the value of the native-only model, 43.6 kcal/mol. When it is combined with the barrier change, we estimate $\Delta F_d \simeq 6\Delta F^{\ddagger}$ for the full Hamiltonian $H(m = 1)$.

What about the formation of traps? Essentially, we are exploring the activation temperature of the spin glass analogy "T_A" for the class of present Hamiltonian. One applies the analogue of the variational scheme for native-state free energy profiles to structures that are not natively folded but, rather, are so-called decoy structures. The only difference for calculating trap escape, as opposed to unfolding of a native structure, is that one must set the $\vec{r}^*$ of the trial Hamiltonian, $H^v = H_B + \sum_j C_j(\vec{r}_j - \vec{r}_j^*)^2$, to be provided by the decoy structure, not to be the coordinates of the native structures, as in the previous calculation.

To measure the stability of trap structures, we again introduce two characteristic temperatures, T_e and T_a. These correspond to the temperatures T_f and T_h for the native structures, respectively. According to the glass analogy,[47,48] T_e is analogous to the ideal glass transition temperature, T_K. Temperature T_a corresponds to the onset of activated trap escape. Their importance has been stressed in the introduction. More specifically, T_e is defined as the temperature when the basin of the trap has the same free energy as the entire unfolded basin. This corresponds to a configurational entropy crisis like that of mean-field spin glass. T_a, on the other hand, is the temperature below which that trap starts to show metastability. At a temperature lower than the activation temperature, T_a, for a particular trap, the trap would emerge as a structure sufficiently long living that it might present some troubles for the Brownian motion of protein to correct, even if the native basin finally is more stable.

We sampled approximately 80 trap structures by AMH simulations and calculated their properties. Some of these trap structures and their contact maps are displayed in Figure 10. The characteristic temperatures of these traps, T_e versus T_a, are shown in Figure 11. We see that the ratio of the two critical temperatures for these traps is similar to (but slightly higher than) those found for the free energy profiles for access to the native structure, that is, $T_a/T_e \simeq 1.3$. All T_e of the traps are lower than the T_f of the native ensemble. This is in harmony with the minimal frustration principles. A more interesting presentation of the trap structures' properties is shown in Figure 12. Again, each dot depicts the properties of one particular trap. We see that a trap's T_e depends strongly on its nativeness as measured by the native energy part of the total energy of the trap basin. In the lower right corner, one finds near native structures that have $T_e \simeq T_f$, and their values of E_{nat} are similar to that of native-only calculation. A slight gap, devoid of dots, separates these near native basin structures from others that spread out more in the upper left regions. For a fixed, low value of stability (T_e), the distribution of the nativeness ($-E_{nat}$) is wider. The traps near the upper contour line are traps having very little native content that are, therefore, primarily stabilized by nonnative interactions. Such structures would be of greatest physiological relevance and would be most challenging to de novo structure prediction.

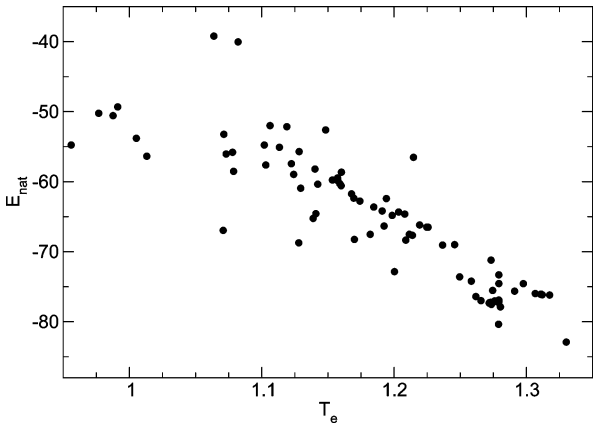

Figure 12. T_e versus E_{nat}, the native energy part of the total energy for a set of (mis)folded basins.

V. Conclusions

Drastic changes of free energy profiles thermodynamically far away from the folding equilibrium are found using variational methods. The present approximation scheme allows one to identify two temperatures characterizing quantitative changes in the folding and unfolding mechanisms. For well-evolved proteins, we find that these two critical temperatures, T_d and T_h, seem to scale with T_f, that is, $T_d \simeq 0.5T_f$ and $T_h \simeq 1.26T_f$, in general. The values of T_d are consistent with earlier analytical results for a idealized model with a generic protein structure[17] and are consistent with simulation results.[49] The scaling holds over a wide range of folds. The way the free energy profile changes with temperatures seems to be quite universal and somewhat independent of detailed folds. We find the additional stabilization free energy required to achieve a downhill landscape at T_d, ΔF_d, is much higher than the free energy barrier present at the equilibrium T_f. An approximate linear relation, $\Delta F_d \simeq 10\Delta F^{\ddagger}$, holds for the few proteins we have calculated, where $\Delta F^{\ddagger}$ is the barrier height of the native-only model. Of course, achieving downhill folding by specific sequence redesign may be easier than this. We also note that ΔF_d is much easier to calculate and thus provides an ideal quick estimation for $\Delta F^{\ddagger}$ via this relation.

We also examine the effects of nonnative interactions on the free energy profiles focusing on the specific case of protein G. At the equilibrium folding temperatures, we find the barrier is increased by a factor of 1.8. The scaling relations between the characteristic temperatures nevertheless seem to be preserved when nonnative interactions are considered. When nonnative interactions are included, the relation between the critical drop of free energy and barrier height is revised to $\Delta F_d \simeq 6\Delta F^{\ddagger}$.

In addition to characterizing profiles for forming native structures, we also studied profiles of forming realistic trap structures. The variational scheme also yields the characteristic temperatures for trap escape, T_e and T_a, for these decoy structures. Like the native structure's counterpart T_f, the temperature T_e provides a useful measure of how stable a trap is. We find a similar linear relation, $T_a \simeq 1.3T_e$, for the onset of activated trap escape.

We studied the trap stabilities and labeled them on the parameter diagram constructed by nativeness versus stability. The high stability structures have a rather narrower variation in nativeness.

We believe some of these features reported here can provide some guidance for simulation or experimental studies of folding when carried out under a larger range of thermodynamic conditions than has been customary.

Acknowledgment. This work was supported by NIH Grant R01 GM044557. The computational resources were provided in part by the NSF-based Center for Theoretical Biological Physics. We thank M. Prentiss for helpful discussions.

Supporting Information Available: Table including values PDB ID, chain length, contact number T_d, T_f, T_h, energy of folded basin at T_h^-, energy at unfolded basin T_d^+, energy of unfolded basin at T_f, energy of folded basin T_f, free energy at T_f, and free energy at T_d^-. This material is available free of charge via the Internet at http:// pubs.acs.org.

References and Notes

(1) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(2) Fersht, A. *Structure And Mechanism In Protein Science: A Guide To Enzyme Catalysis And Protein Folding*; W. H. Freeman: New York, 1999.
(3) Frauenfelder, H., Deisenhofer, J., Wolynes, P. G., Eds. *Simplicity and Complexity in Proteins and Nucleic Acids*; Dahlem University Press: Berlin, 2000.
(4) Oliveberg, M.; Wolynes, P. G. *Q. Rev. Biophys.* **2005**, *38*, 245−288.
(5) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Hanser-Gardner: Munich, 1989.
(6) Doi, M.; Edwards, S. F. *The Theory of Polymer Dynamics*, 2nd ed.; Oxford University Press: Oxford, 1986.
(7) de Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(8) Naganathan, A. N.; Doshi, U.; Munoz, V. *J. Am. Chem. Soc.* **2007**, *129*, 5673−5682.
(9) Ferguson, N.; Schartau, P. J.; Sharpe, T. D.; Sato, S.; Fersht, A. R. *J. Mol. Biol.* **2004**, *344*, 295−301.
(10) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(11) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* **2000**, *41*, 415−427.
(12) Bross, P.; Corydon, T. J.; Andresen, B. S.; Jorgensen, M. M.; Bolund, L.; Gregersen, N. *Hum. Mutat.* **1999**, *14*, 186−198.
(13) Selkoe, D. J. *Nature* **2003**, *426*, 900−904.
(14) Portman, J. J.; Takada, S.; Wolynes, P. G. *Phys. Rev. Lett.* **1998**, *81*, 5237−5240.
(15) Portman, J. J.; Takada, S.; Wolynes, P. G. *J. Chem. Phys.* **2001**, *114*, 5069−5081.
(16) Portman, J. J.; Takada, S.; Wolynes, P. G. *J. Chem. Phys.* **2001**, *114*, 5082−5096.
(17) Takada, S.; Wolynes, P. G. *J. Chem. Phys.* **1997**, *107*, 9585−9598.
(18) Lee, C.-L.; Lin, C.-T.; Stell, G.; Wang, J. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2003**, *67*, 041905.
(19) Leite, V. B.; Onuchic, J.; Stell, G.; Wang, J. *Biophys. J.* **2004**, *87*, 3633−3641.
(20) Wang, J. *Biophys. J.* **2004**, *87*, 2164−2171.
(21) Bryngelson, J. D.; Onuchic, J.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct. Funct. Genet.* **1995**, *21*, 167−195.
(22) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524−7528.
(23) Wang, J.; Plotkin, S. S.; Wolynes, P. G. *J. Phys. I France* **1997**, *7*, 395−421.
(24) Goldstein, R. A.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4918−4922.
(25) Prentiss, M. C.; Hardin, C.; Eastwood, M. P.; Zong, C.; Wolynes, P. *J. Chem. Theory Comput.* **2006**, 2, 705−716.
(26) Sasai, M.; Wolynes, P. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1992**, *46*, 7979−7997.
(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(28) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. *Biochemistry* **1994**, *33*, 4721−4729.
(29) Qi, X.; Portman, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10841−10846.
(30) Shen, T.; Hofmann, C. P.; Oliveberg, M.; Wolynes, P. G. *Biochemistry* **2005**, *44*, 6433−6439.
(31) Zong, C.; Wilson, C. J.; Shen, T.; Wolynes, P. G.; Wittung-Stafshede, P. *Biochemistry* **2006**, *45*, 6458−6466.
(32) Stoessel, J. P.; Wolynes, P. G. *J. Chem. Phys.* **1984**, *80*, 4502−12.
(33) Shen, T.; Wolynes, P. G. *New J. Phys.* **2006**, *8*, 273.
(34) Shen, T.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8547−8550.
(35) Bixon, M.; Zwanzig, R. *J. Chem. Phys.* **1978**, *68*, 1896−1902.
(36) Miyazawa, S.; Jernigan, R. *J. Mol. Biol.* **1996**, *256*, 623−644.
(37) Hardin, C.; Eastwood, M. P.; Prentiss, M. C.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1679−1684.
(38) Zong, C.; Papoian, G. A.; Ulander, J.; Wolynes, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 5168−5176.
(39) Hardin, C.; Eastwood, M.; Luthey-Schulten, Z.; Wolynes, P. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14235−14240.
(40) Hardin, C.; Eastwood, M. P.; Luthey-Schulten, Z.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14235−14240.
(41) Eastwood, M. P.; Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. *IBM J. Res. Dev.* **2001**, *45*, 475−497.
(42) Papoian, G. A.; Ulander, J.; Eastwood, M. P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 3352−3357.
(43) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6170−6175.
(44) Plotkin, S. *Proteins* **2001**, *45*, 337−345.
(45) Cieplak, M.; Hoang, T. *Int. J. Mod. Phys. C* **2002**, *13*, 1231−1242.
(46) Clementi, C.; Plotkin, S. S. *Protein Sci.* **2004**, *13*, 1750−1766.
(47) Gotze, W. *J. Phys. Condens. Matter* **1999**, *11*, A1−A45.
(48) Kirkpatrick, T. R.; Wolynes, P. G. *Phys. Rev. A: At., Mol., Opt. Phys.* **1987**, *35*, 3072−3080.
(49) Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. *Proteins* **1999**, *34*, 281−294.