Quantifying Protein Folding Transition States with Φ_T

J. ERVIN¹ and M. GRUEBELE^{1,2,*}

Departments of Chemistry,¹ Physics,² and Biophysics and Computational Biology², University of Illinois at Urbana-Champaign, IL 61801, USA

(*Author for correspondence, E-mail: gruebele@acs.uiuc.edu)

Abstract. A number of reaction coordinates have been proposed for reduced-dimensionality representations of a protein's folding free energy surface. We discuss in detail the entropic reaction coordinate $\Phi_T = \Delta S^\ddagger/\Delta S$, recently introduced to quantify the conservation of mutations and the location of the folding transition state based on experimental temperature-tuning data. Numerical simulations illustrate the advantages as well as the limitations of Φ_T . Φ_T can be determined from experiment, computation, and analytical theory; Φ_T can also be used to investigate structurally localized perturbations of the free energy surface. However, Φ_T is only a relative reaction coordinate; furthermore, proteins undergo cold denaturation at sufficiently low temperatures, and care must be taken in interpreting Φ_T near the region where $\partial\Delta G/\partial T = 0$, particularly if the heat capacity change upon folding is small.

Key words: free energy, Kramers model, phi-value analysis, reaction coordinate, ubiquitin

1. Introduction

The mere fact that proteins fold at all is a sufficient condition for the existence of a small number of reaction coordinates, and this is embodied in the principles of consistency and minimal frustration [1, 2]. Yet fast protein folding processes at early times are so heterogeneous that a single global coordinate will not likely describe them in full. Even if one could transform the 2–5 dihedral angles effectively available to each residue to a serviceably small set of coordinates, one would *prima facie* neglect complex solvation effects [3]. Although small molecules are well described by a single reaction coordinate, and macroscopic phase transitions are well described by a single order parameter, the mesoscopic nature of protein folding forces theorist and experimentalist alike to consider a whole zoology of folding coordinates.

Still, all is not lost. The combination of several global reaction coordinates has led to a fruitful multi-dimensional dissection of the folding free energy surface, and many correlated effects have been uncovered by simulation and analytical theory [4–8]. Experiments have also measured several different reaction coordinates, and much can be learned by how ‘independent’ these reaction coordinates are [9–11].

In some cases, it has been shown explicitly that a single global coordinate consists of a changing set of local coordinates, whose composition depends on protein sequence and solvent conditions [12].

To bring theoretical and experimental analyses together, it is useful to have at one's disposal a set of reaction coordinates which can be measured as well as computed, so the folding free energy surface can be examined in terms of all of these coordinates. In addition, it would be useful to have a set which is sensitive to local structural variations, so the powerful tools of protein engineering [13, 14] can be exploited.

Many commonly used folding coordinates do not satisfy all of these criteria. For example, the native contact fraction Q , which has proved very useful in both simulations and analytical folding theories [4], cannot be determined experimentally. This is partly due to experimental difficulties (e.g. NOE peaks cannot be measured with sufficient time resolution), and partly due to inherent ambiguities in the definition of a native contact. As another example, the radius of gyration R_g can be measured and computed [6], but R_g is not particularly sensitive to local variations in structure and free energy (resulting from single point mutations, for example).

One class of folding coordinates does satisfy the requirements laid out above: those based on thermodynamic tuning. They rely on the postulate set forth by G.S. Hammond [15], that a given reactive ensemble tends to be structurally intermediate between nearby ensembles surrounding it on the reaction coordinate; consequently its free energy tunes by an intermediate amount under externally applied stresses. For proteins, the caveat implies that in a multidimensional reaction coordinate space, closeness along a single reaction coordinate does not imply closeness along other reaction coordinates. Nonetheless, by studying different thermodynamic stresses, one can map out regions of the free energy surface and assign structural correlations. Let the free energy depend on a 'stress' or 'tuning' parameter c (c = temperature, pressure, pH, denaturant, etc.). Then $\partial G^\dagger(c)/\partial c$ will generally lie between $\partial G_U(c)/\partial c$ and $\partial G_F(c)/\partial c$ if the state $\dagger$ lies between the states U (denatured) and F (folded). Referencing all free energies to the state U , it follows that

$$0 \leq \Phi_c = \frac{\partial \Delta G^\dagger(c)/\partial c}{\partial \Delta G_F(c)/\partial c} \leq 1 \quad (1)$$

as long as the Hammond postulate is obeyed along the projected reaction coordinate conjugate to c , as defined by eq. (1). The middle of eq. (1) then defines a reaction coordinate or order parameter Φ_c which tunes from 0 to 1 depending on whether the state $\dagger$ lies close to U or F .

This approach has been used fruitfully since the mid-90s by tuning protein free energies with denaturants such as urea or guanidinium hydrochloride. Initial studies verified $\Phi_{[d]} \approx 0.5$ for cytochrome c [9], and energy landscape theory predicted that curvature would be seen in $\Phi_{[d]}$, moving it beyond a linear free

energy model. Such correlations, corresponding to motion of the transition state, were indeed uncovered [10, 16]. Recently, other Φ_c , notably Φ_T [11], $\Phi_{[salt]}$ and Φ_{pH} [17] have been studied, and it has been pointed out how several of these in combination with site-directed mutagenesis could be used to probe local structure formation and local reaction coordinates [12]. Furthermore, large shifts in Φ_c upon mutationn can be used to flag nonconservative mutations which move the transition state.

The main limitation of coordinates derived by thermodynamic tuning is that they are relative, not absolute, reaction coordinates. For instance, a decrease of Φ_c of the transition state could arise because the transition state becomes less native-like or because the unfolded state becomes more native like. This is a nontrivial piece of missing information: for example, it determines what folding ensembles can be reached in a relaxation experiment where the thermodynamic parameter c is suddenly changed [18]. Another potential limitation, discussed in detail later, is that eq. (1) could be nonmonotonic, or even discontinuous in c over the full interval of interest.

In this paper, we study the coordinate Φ_T . One of the systems most extensively studied by site-directed mutagenesis has also been studied systematically using Φ_T [11, 12]. Φ_T serves as a reminder that eq. (1) is not always obeyed: proteins suffer a reentrant phase transition at low temperature ('cold denaturation') [19–21], which causes discontinuities and sign changes in Φ_T . However, we shall see that judicious scaling of Φ_T (or other reaction coordinates obtained through thermodynamic tuning) can be used to eliminate these discontinuities leading to coordinates useful over the whole range of the parameter c (here T).

2. The Φ_T Coordinate: Simulation

The free energy of folding of a solvated protein depends on a set of thermodynamic 'stresses' or 'tuning parameters' $\mathbf{c} = \{c_i\}$. In the Gibbs ensemble, natural parameters include temperature T , pressure P , and concentrations M_i of the various solution components: $\Delta G = \Delta G(\mathbf{c})$. Within the transition state approximation, we can also write the free energy of activation as a function of these parameters: $\Delta G^\ddagger = \Delta G^\ddagger(\mathbf{c})$.

The free energies do not necessarily depend monotonically on $\mathbf{c}$. For example, proteins cold denature and heat denature, and therefore $\Delta G(\mathbf{c})$ must be at least a quadratic function of the temperature:

$$\Delta G(T - T_0) = \Delta G_0 + \Delta G_2(T - T_0)^2 + \dots \quad (2a)$$

$$\Delta G^\ddagger(T - T_0^\ddagger) = \Delta G_0^\ddagger + \Delta G_2^\ddagger(T - T_0^\ddagger)^2 + \dots \quad (2b)$$

Without loss of generality, we can choose $\Delta G_1 = 0$ by letting T_0 be the temperature at the free energy minimum ΔG_0 . The activation free energy in eq. (2b) is expanded in the same form as the free energy of folding. Quadratic terms (i.e.

cold denaturation-like behavior) have indeed been observed for intermediates and transition states [12, 20, 22], and in some cases $T_0^\dagger \neq T_0$ has been determined.

In the presence of other tuning parameters, we can still choose the minimum of ΔG at $\mathbf{c} = 0$, so the multidimensional version of eq. (2) remains quadratic to lowest order:

$$\Delta G(\mathbf{c}) = \Delta G_0 + (\mathbf{c} - \mathbf{c}_0)^T \cdot \Delta G_2 \cdot (\mathbf{c} - \mathbf{c}_0) + \dots \quad (3a)$$

$$\Delta G^\dagger(\mathbf{c}) = \Delta G_0^\dagger + (\mathbf{c} - \mathbf{c}_0^\dagger)^T \cdot \Delta G_2^\dagger \cdot (\mathbf{c} - \mathbf{c}_0^\dagger) + \dots \quad (3b)$$

where T indicates the vector transpose and ΔG_2 is a matrix of coefficients. For simplicity we will use the Taylor expansion in eqs. (2–3). More physically motivated expansions starting with a heat capacity change ΔC_p [21], as well as other variables, are perfectly reasonable choices also. Most importantly for the present purpose, the quadratic form in eqs. (2–3) are the minimum required to describe the simultaneous cold and heat denaturation of proteins.

Although the free energies are not globally monotonic function of $\mathbf{c}$, they may be locally monotonic. For example, $\partial \Delta G / \partial M_{GuHCl} \geq 0$ for reasonably high concentrations of guanidinium hydrochloride denaturant, and $\partial \Delta G / \partial T = -\Delta S \geq 0$ near the midpoint T_m of the heat denaturation transition. Over such a range, we can construct reaction coordinates Φ_{c_i} by defining

$$\Phi_{c_i} = \frac{\partial \Delta G^\dagger(\mathbf{c}) / \partial c_i}{\partial \Delta G(\mathbf{c}) / \partial c_i} = \frac{\mu_i^\dagger(\mathbf{c})}{\mu_i(\mathbf{c})}, \quad (4)$$

where μ_I is the chemical potential conjugate to thermodynamic variable c_i . Eq. (2) satisfies the basic requirements for a reaction coordinate used to locate a reactive intermediate ensemble † between the reactant (denatured protein U) and product (native protein F): The free energies of the folded state ΔG and of the transition state $\Delta G^\dagger$ are referenced to the unfolded state, and because the thermodynamic properties of the transition state are intermediate between the unfolded and folded state when the Hammond postulate is valid, Φ_{c_i} lies between 0 and 1.

We are mainly interested here in temperature as a tuning parameter, so

$$\Phi_T(T) = \frac{\partial \Delta G^\dagger / \partial T}{\partial \Delta G / \partial T} = \frac{\Delta S^\dagger(T)}{\Delta S(T)} \quad (5)$$

Eq. (5) describes how closely the entropy change of the ensemble resembles the entropy change achieved in the fully folded state. For simplicity, we will illustrate eq. (5) with an expansion of ΔG up to quadratic terms only (eq. (2a)). This often suffices to describe experiments quantitatively (see section 3). The expansion of $\Delta G^\dagger$ involves several more subtle issues. First, $\Delta G^\dagger$ must contain higher order terms than ΔG , or $T_0^\dagger \neq T_0$; otherwise Φ_T would be a constant reaction coordinate. Furthermore, the expansion coefficients must be chosen such that $\Delta G^\dagger > \Delta G$ at all temperatures where a transition state theory is to be applied. Finally, at high

temperatures folding is not favored, the transition state lies late along the reaction coordinate, and $\partial\Delta G^\ddagger/\partial T$ approaches $\partial\Delta G/\partial T$ according to the Hammond postulate.

These physically motivated constraints can be imposed by expanding $\Delta G^\ddagger$ nominally to same order as ΔG , but making $\Delta G_2^\ddagger$ in eq. (2b) temperature-dependent in the following fashion:

$$\Delta G_2^\ddagger = \Delta G_2 + \Delta\Delta G_2^\ddagger \frac{\Delta T^2}{\Delta T^2 + (T - T_0^\ddagger)^2}. \quad (6)$$

The Lorentzian interpolation function assures that under high thermal stress, $\Delta G_2^\ddagger$ approaches ΔG_2 , so that $\Delta G^\ddagger \geq \Delta G$ if ΔT is not chosen too large. Under low thermal stress, $\Delta G^\ddagger$ approaches $\Delta G_2 + \Delta\Delta G_2^\ddagger$. The functional form of the interpolation function is immaterial as long as it does not decay too fast. (For example, a Gaussian would produce an additional inflection point in $\Delta S^\ddagger$, which is undesirable.) Eq. (6) can be extended to higher orders n , as long as the higher order terms are defined in analogy to eq. (6). For each order n , there would be a $\Delta\Delta G_n^\ddagger$ and a ΔT_n as additional parameters. Table I shows a reasonable set of parameters in reduced units, which are chosen such that when interpreted in $^\circ\text{C}$ (temperatures) and kJ/mole (free energies), they correspond closely to values for real proteins. In particular, $\Delta\Delta G_2^\ddagger < 0$ and $\Delta T = 40$ corresponds to an early transition state under favorable folding conditions, which becomes late once the midpoint of the thermal denaturation curve is passed. This is typical of the behavior observed in real proteins [11].

Figure 1 (left) shows the case when $T_0^\ddagger - T_0 = 0$. In that case eq. (5) is singular only at $T - T_0 = 0$. The thin solid line in Figure 1 (left) shows the reaction coordinate Φ_T , which is a minimum where the transition state has its lowest free energy, and increases both under cold and heat denaturation. This behavior is expected because eq. (5) is based on the van 't Hoff entropy. Φ_T can be used as a reaction coordinate for both cold and heat denaturation. It increases (later transition state) when the protein is stressed in either direction from the optimal folding temperature. The fact that Φ_T is nonmonotonic near $T - T_0 = 0$ is simply

Table I. Reduced thermodynamic and kinetic parameters for the illustrative simulations in Figures 1 and 2. The values are chosen so that if free energies are expressed in kJ/mole and temperatures in $^\circ\text{C}$, free energies representative of real proteins are obtained. (cf also the values for ubiquitin in Table II)

$T_0 = 0$	$T_0 - T_0^\ddagger = 0, -4, 4 \text{ or } 8 \text{ in Figures 1 and 2}$
$\Delta G_0 = -15$	$\Delta G_0^\ddagger = +10$
$\Delta G_2 = 0.005$	$\Delta\Delta G_2^\ddagger = -0.004$
$\Delta G_3' = 0 \text{ or } 0.00001$	$\Delta T = 100 (\infty \text{ in Figure 2 middle})$

Figure 1. Free energies ΔG and $\Delta G^\ddagger$ (thick solid curves), folding coordinate Φ_T (dotted curve), scaling factor Φ_T' (dashed curve) and scaled folding coordinate $\bar{\Phi}_T$ (thin solid curve). Left: native state and transition state have equal temperatures of maximal stability. Middle and right: temperatures of maximal stability are shifted by 4 degrees.

Figure 2. Free energies ΔG and $\Delta G^\ddagger$ (thick solid curves), folding coordinate Φ_T (dotted curve), scaling factor Φ_T' (dashed curve) and scaled folding coordinate $\bar{\Phi}_T$ (thin solid curve). Left: temperatures of maximal stability for the native and transition states are shifted by 8 °C. Middle: Large ΔT causes the activation energy to vanish above a limiting temperature. Right: higher order terms can be included in the analysis, here a cubic term from Table I.

an indication that the transition state ensembles surrounding the minimum free energy state entropically resemble each other more than either resembles the lowest free energy state. The free energies still satisfy the Hammond postulate: the states surrounding the optimal state are similar to the optimal state.

The situation is more complicated when $T_0^\ddagger - T_0 \neq 0$ (Figure 1 middle and right). In that case the denominator of eq. (5) introduces a singularity. This leads to a change in Φ_T even when $\Delta\Delta G_2 = 0$ and when the curvatures of the free energy and of free energy of activation surfaces are identical (dotted lines). Clearly this $T_0 - T_0^\ddagger$ shift can cause Φ_T to move well outside of the (0,1) range. Figure 2 (left) illustrates similar behavior when $T_0 - T_0^\ddagger$ is twice as large as in Figure 1, showing the linear scaling of the singularity. The region where Φ_T is not useful as a reaction coordinate extends by about $2|T_0^\ddagger - T_0|$ on either side of the points of maximal stability of ΔG and $\Delta G^\ddagger$. This generally has no significant effect near the heat denaturation midpoint, but would hamper the use of Φ_T as a reaction coordinate

across the full temperature range. In particular, if the heat capacity change upon folding is small and $T_0 - T_0^\dagger$ is large, negative Φ_T values could result, indicating non-native behavior.

The contribution of this singularity to the reaction coordinate can be removed by introducing a hybrid free energy

$$\Delta G' = \Delta G_0 + \Delta G_2(T - T_0^\dagger)^2 + \dots \quad (7)$$

It is identical to eq. (2b) except that all thermodynamic parameters of the intermediate ensemble save the temperature of maximum stability have been replaced by those of the native state. In analogy to eq. (5) we can define a ‘trivial’ normalization coordinate

$$\Phi'_T(T) = \frac{\partial \Delta G' / \partial T}{\partial \Delta G / \partial T} \quad (8)$$

normalization coordinate has the same singularity as Φ_T , but rapidly approaches unity when the thermodynamic stress is increased. The scaled coordinate

$$\overline{\Phi}_T = \Phi_T / \Phi'_T \quad (9)$$

therefore closely matches Φ_T near the midpoint of the heat denaturation transition, but restores the desired behavior of a global reaction coordinate which accounts for both heat and cold denaturation near $T - T_0 = 0$. This continuous coordinate is indicated by the thin black curve in Figures 1 and 2.

Figure 2 (right) shows that the same concept can be applied to higher order expansions of the free energy. Consistent inclusion of cubic terms in the free energies of folding and activation again leads to a smooth coordinate $\overline{\Phi}_T$ which agrees with Φ_T everywhere except near the free energy minima. Finally, it should be reiterated once more that even eq. (6) can lead to unphysical behavior, if ΔT is chosen so large that the intermediate ensemble never approaches the curvature of the native state at elevated temperature (Figure 2 middle).

3. The Φ_c coordinates: application to experiment

The great advantage of eqs. (5) and (9) is that they are accessible by experiment (requiring rate and equilibrium measurements if Φ is to be determined for the transition state ensemble) and also by analytical or computational models. Among the latter, free energies can be deduced from a reasonable Ansatz for the protein partition function, or simulated in reduced or all-atom models. Full-atom molecular dynamics simulations of free energies are still rather expensive (requiring 100s of ns of trajectories) [23]. Recently, methods for evaluating the necessary entropies (or free energy differences) efficiently have been discussed in the literature [24, 25], and off-lattice models can easily evaluate the required quantities already.

The generalization to more coordinates is obvious, and in fact chaotropic denaturant tuning has been used for years to study moving reaction coordinates [9,

16]. The advantage of variables such as T or P is that the conjugate quantities (ΔS or ΔV) are exactly defined and straightforward to compute. Chaotrope concentration is related to exposed hydrophobic surface area [26], but this is only a correlation. The differential solvation layers which occur upon chaotropic denaturation make a fully quantitative interpretation more difficult. Nonetheless, tuning urea, salt, pH or other solute parameters M_i provide useful parameters complementary to temperature and pressure.

Eq. (5) has long been applied to discrete variables c such as mutations. The discrete version of eq. (5) underlies the technique of protein engineering developed by Fersht and coworkers, and yields a Φ_m value [13]. A potential problem with discrete changes is that they cannot be made infinitesimal, and so they may unduly perturb the intermediate reactive ensemble. It was recently shown how protein engineering and continuous thermodynamic variables can be combined to classify mutations as ‘conservative’ or ‘non-conservative,’ and how information about the local energy landscape other than the transition state can be extracted from ‘nonconservative’ mutations [12]. Basically, nonconservative mutations can cause a sudden shift in ϕ_T , while conservative mutations do not affect the entropic character of the transition state. Ideally several ϕ_c should be used to verify conservatism of a mutation, whenever large heat capacity changes upon mutation make ϕ_T problematic, the mutation is of course already proved non-conservative.

Multidimensional tuning of several variables will probe how valid the Hammond postulate is for proteins. The extent of the postulate’s validity for any single reaction coordinate is deeply connected to the question of how many independent reaction coordinates are required to describe fully the folding mechanism(s) of a given fold over a broad space of solvent conditions and sequences. Cross terms in the matrix ΔG_2 in eq. (3) and the independence of thermodynamic folding coordinates (or lack thereof) defined by eq. (4) will allow this question to be probed by experiments. Simulations already have made use of the idea of constructing a small number of coordinates embedded in a higherdimensional space of coordinates such that a minimal number of such coordinates accounts for as much of the free energy surface as possible [5].

Finally, it should be noted that although the discussion in this section focused mainly on determining Φ_T for the *transition state* ensemble, it applies to the characterization of any intermediate ensemble. It is particularly easy to measure Φ_T for low-energy thermodynamic intermediates: their free energy can be determined reliably because they are separated by barriers from the folded and unfolded states, and can be observed under steady-state conditions. A greater challenge is offered by high energy intermediates, or reactive ensembles ‘hidden’ kinetically by a dominant transition state. Recent progress in that direction has been made by combining protein engineering with Φ_T analysis. For the WW domain, it was found that once the residues making native-like contributions to the free energy of the transition state were identified, other residues could be probed as follows: if mutations of increasing severity are made at a site which does not limit the rate in the wild

Figure 3. Experimental results for the ubiquitin Trp45Phe mutant Ub* in the region of $T_0^\ddagger$ and T_0 . The thermodynamic fit of the integrated fluorescence data in A is to a two state model with quadratic free energy. The kinetic fit to the refolding rate is to a Kramers model with ‘normal’ viscosity dependence. These two fits yield $\Delta G(T)$ and $\Delta G^\ddagger(T)$. Free energy derivatives should be calculated from such fits, not from the raw data.

type protein, it eventually becomes rate-limiting and Φ_T suddenly increases. This indicates that a switchover to a new transition state has occurred, and the data can be fitted to extrapolate the local free energy at the site in the wild-type [12]. Thus, Φ_T can be used to identify ‘hidden’ regions of the free energy surface, and to quantify when a mutation becomes ‘nonconservative’ (i.e. shifts the transition state for folding). Such studies will become increasingly powerful as more reaction coordinates are added.

4. The ϕ_T coordinate: comparison with experiment

Φ_T has already been used to study folding reactions near the heat denaturation transition. The experimental challenges for obtaining ϕ_T are similar to those in mutant phi-value analysis. However, to obtain ϕ_m raw data must be subtracted. ϕ_T has the advantage of being a continuous function of T . As seen in Figure 3B, it would be undesirable to obtain $\delta\Delta G^\ddagger/\delta T$ from raw data. Fortunately, if many rates are measured, a smooth function such as eq. 2 can be used to fit the data (Figure 3B), yielding reliable free energy derivatives. Over a narrow temperature range near T_m , quadratic fits with different T_0 and $T_0^\ddagger$ are often sufficient [11, 12].

Here, we are interested in the more global properties, and in particular in the singular region near T_0 and $T_0^\ddagger$. In order to study this, we performed measurements on a ubiquitin mutant under acidic conditions where both cold and heat denaturation could be observed.

Table II. Thermodynamic and kinetic fitting parameters for ubiquitin, from fits to the data in Figure 3. The same free energy model as for Table I was used

$T_0 = 21.3\text{ }^\circ\text{C}$	$T_0; T_0^\ddagger = 7.2\text{ }^\circ\text{C}$
$\Delta G_0 = -6.1(2)\text{ kJ/mole}$	$\Delta G_0^\ddagger = 40.3(5)\text{ kJ/mole}$
$\Delta G_2 = 0.0071(2)\text{ kJ/mole-K}^2$	$\Delta \Delta G_2^\ddagger = -0.0015(3)\text{ kJ/mole-K}^2$
	$\Delta T = 40\text{ K (fixed)}$

The Phe45Trp mutant of human ubiquitin was generated as described in refs. [27, 28] (using an expression system obtained from Tracy Handel [29]). Measurements were carried out in a 0.75 M GuHCl solution at pH 2.3. The temperature titration was monitored by fluorescence excited at 280 nm and integrated at wavelengths above 320 nm, at a protein concentration of 5 μM (Figure 3A). Stopped-flow data to measure the relaxation rate was taken by diluting a 1.5 M urea, pH 2.3 50 μM protein solution 1:1 with 0 M urea, pH 2.3 refolding buffer, so kinetics could be monitored under the same solution conditions as the steady state data. Fluorescence was likewise excited at 280 and collected above 320 nm. Kinetic traces were averaged 5–15 times and fitted to a double exponential, with the first, faster exponential consistently representing >80% of the amplitude. No significant burst phase was resolvable using fluorescence detection. A very fast burst phase due to downhill formation of a more structured denatured state, has been detected by time-resolved X-ray scattering and circular dichroism [28]. This initial relaxation is so fast that it would not affect the treatment of the folding phase observed by fluorescence as two state folding. The large fluorescence phase is therefore assigned to formation of the native state over a single barrier, and the small slower phase to a combination of prolyl isomerization events which have been studied in detail elsewhere [27, 30].

The steady state data were analyzed by a two state model using eq. (2) to compute native and denatured protein concentrations as a function of temperature. A linear expansion of the fluorescence intensities of the native and denatured states as a function of temperature was used:

$$\begin{aligned}
 Fl_{obs} &= Fl_U[U] + Fl_F[F] \\
 Fl_U &= Fl_{U0} + Fl_{U1}(T - T_0) \\
 Fl_F &= Fl_{F0} + Fl_{F1}(T - T_0)
 \end{aligned}
 \tag{10}$$

The least-squares fit of this model to the data shown in Figure 3A is highly accurate, showing that the steady-state fluorescence data are entirely consistent with a two state model. They are also consistent with NMR data on ubiquitin mutants of equal stability which show that only 85–90% native population is reached at the optimal folding temperature [17].

Figure 4. Thermodynamic folding coordinates obtained from the fits in Figure 3. (A) The discontinuity in Φ_T and its correction by scaling are shown in the left panel. The thin solid curve shows that at pH 2.3 and 0.75 M GuHCl, Ub* has a late transition state whose position depends only slightly on temperature. (B) Shows the difference between eq. (6) and a purely quadratic model with T-independent $\Delta\Delta G_2^\ddagger$, extrapolated to higher temperatures. Folding data must be acquired over the widest possible range above T_m to yield a global fit which also maps out transition state motion.

The folding rate constant $k_f(T)$ was obtained from $k_{obs} = k_f(1+K)/K$ by using the equilibrium constant $K(T)$ obtained from the temperature titration. The kinetic data were analyzed to obtain $\Delta G^\ddagger(T)$ by using a Kramers model [13] with normal (inverse) viscosity dependence, and a calculated viscosity of the aqueous GuHCl solution as a function of temperature [32]:

$$k_f(T) = v^\ddagger \frac{\eta_0}{\eta(T)} e^{-\Delta G^\ddagger(T)/RT} \quad (12)$$

Beyond the viscosity, the choice of the multiplicative constant frequency prefactor $v^\ddagger$ is somewhat ambiguous [33, 34], but of no concern here because it does not contribute to the derivatives necessary to evaluate Φ . The parameters obtained from all the thermodynamic and kinetic fits are summarized in Table II. Figure 3B shows the model fit to the refolding rate constant.

Figure 4A shows the reaction coordinates Φ_T and $\overline{\Phi}_T$ obtained from the thermodynamic parameters in Table II. Φ_T shows the expected discontinuity at T_0 , which is removed by $\overline{\Phi}_T = \Phi_T / \Phi_T'$. In the case of ubiquitin, the correction to Φ_T by Φ_T' is large: Φ_T actually lies above both $\overline{\Phi}_T$ and Φ_T' . The large value of $\overline{\Phi}_T$ indicates that the ubiquitin transition state is rather late under the present conditions (0.75 M GuHCl, pH 2.3), in keeping with the fact that the denaturant and low pH bias the folding conditions away from the native state. Destabilization of the product should result in a late transition state. A small variation in $\overline{\Phi}_T$ is discernible when ΔT is fixed at 40 °C: the transition state becomes more native-like when the conditions for folding are optimized near 25 °C. Recently, more favorable solvent conditions

(e.g. in ethylene glycol solution) have shown that Φ_T can be shifted below 0.1. [28].

Figure 4B illustrates that an even broader range of temperature conditions would be needed to fit ΔT reliably. Significant deviations between a quadratic $\Delta G^\ddagger$ and eq. (6) do not occur until temperatures above 70 °C. A temperature range only near T_0 cannot determine transition state movements very accurately. This problem does not exist when fitting data around T_m , the heat denaturation midpoint. T_m lies sufficiently far from T_0 and $T_0^\ddagger$ that the singularity has a much smaller effect on the Φ_T value (Figure 2 left). Hence a quadratic fit about T_m will produce a different set of thermodynamic parameters, which describe the reaction coordinate near T_m adequately. The global analysis described here will become increasingly useful as more proteins are studied over wide temperature ranges. Already, apomyoglobin, λ -repressor fragment, barstar, ubiquitin and several other proteins have been studied over a sufficiently wide temperature range to require the higher order analysis, and the rank of such proteins will grow in the next few years.

5. Conclusions

Relative protein folding coordinates, obtained by thermodynamic tuning parameters such as T or P, can be computed as well as measured because they are based on well-defined thermodynamic derivatives. Φ_T is a particularly useful coordinate, based on the entropy deficit of intermediate ensembles compared to the native state. It can be used as an order parameter or to reveal nonconservative mutations. A reentrant phase transition such as cold denaturation complicates the picture, but this complication can be removed by suitable scaling to obtain $\overline{\Phi}_T$. $\overline{\Phi}_T$ is monotonic upon changing the temperature away from the point of maximum stability, either in the direction of cold denaturation, or during heat denaturation. Comparison of a number of such coordinates based on T, P denaturant concentration, and other thermodynamic variables, will enable testing of the Hammond postulate for protein folding, and may uncover the need for multiple reaction coordinates.

Acknowledgements

This work was supported by grant PHS RO1 GM057175 of the National Institutes of Health.

References

1. Go, N.: Theoretical Studies of Protein Folding, *Ann. Rev. Biophys. Bioeng.* **12** (1983), 183–210.
2. Bryngelson, J.D. and Wolynes, P.G.: Spin Glasses and the Statistical Mechanics of Protein Folding, *Proc. Natl. Acad. Sci. USA* **84** (1987), 7524–7528.
3. Bolhuis, P.G., Dellago, C. and Chandler, D.: Reaction Coordinates of Biomolecular Isomerization, *Proc. Nat. Acad. Sci. USA* **97** (2000), 5877–5882.

4. Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z. and Wolynes, P.G.: Protein Folding Funnels: The Nature of The Transition State Ensemble, *Folding Des.* **1** (1996), 441–450.
5. Becker, O.M.: Quantitative Visualization of a Macromolecular Potential Energy Funnel, *Theo. Chem.* **398–399** (1997), 507–516.
6. Boczeko, E.M. and Brooks, C.L.I.: First-Principles Calculation of the Folding Free Energy of a Three-helix Bundle Protein, *Science* **269** (1995), 393–396.
7. Socci, N.D., Onuchic, J.N. and Wolynes, P.G.: Diffusive Dynamics of the Reaction Coordinate for Protein Folding Funnels, *J. Chem. Phys.* **104** (1996), 5860–5868.
8. Portman, J.J., Takada, S. and Wolynes, P.G.: Microscopic Theory of Protein Folding Rates. I. Fine Structure of the Free Energy Profile and Folding Routes from a Variational Approach, *J. Chem. Phys.* **114** (2001), 5069–5081.
9. Mines, G.A., Pascher, T., Lee, S.C., Winkler, J.R. and Gray, H.B.: Cytochrome c Folding Triggered by Electron Transfer, *Chem. Bio.* **3** (1996), 491–497.
10. Oliveberg, M.: Alternative Explanations for Multistate, Kinetics in Protein Folding: Transient Aggregation and Changing Transition-State Ensembles, *Acc. Chem. Res.* **31** (1998), 765–772.
11. Crane, J.C., Koepf, E.K., Kelly, J.W. and Gruebele, M.: Mapping the Transition State of the WW Domain Beta Sheet, *J. Mol. Biol.* **298** (2000), 283–292.
12. Jäger, M., Nguyen, H., Crane, J., Kelly, J. and Gruebele, M.: The Folding Mechanism of a β -Sheet: The WW Domain, *J. Mol. Biol.* **311** (2001), 373–393.
13. Fersht, A.R., Matouschek, A. and Serrano, L.: The Folding of an Enzyme I. Theory of Protein Engineering Analysis of Stability and Pathway of Protein Folding, *J. Mol. Biol.* **224** (1992), 771–782.
14. Oliveberg, M., Tan, Y.-J., Silow, M. and Fersht, A.R.: The Changing Nature of the Protein Folding Transition State: Implications for the Shape of the Free Energy Profile for Folding, *J. Mol. Biol.* **277** (1998), 933–943.
15. Hammond, G.S.: A Correlation of Reaction Rates, *J. Am. Chem. Soc.* **77** (1955), 334–338.
16. Ternstrom, T., Mayor, U., Akke, M. and Oliveberg, M.: From Snapshot to Movie: Phi Analysis of Protein Folding Transition States Taken One Step Further, *Proc. Natl. Acad. Sci. USA* **96** (1999), 14854–14859.
17. Ervin, J.: Early Events in the Folding of the Protein Ubiquitin, Ph.D. Thesis, University of Illinois, 2001.
18. Sabelko, J., Ervin, J. and Gruebele, M.: Observation of Strange Kinetics in Protein Folding, *Proc. Nat. Acad. Sci. USA* **96** (1999), 6031–6036.
19. Privalov, P.L.: Cold Denaturation of Proteins, *Crit. Revs. Biochem. Mol. Bio.* **25** (1990), 281–305.
20. Sabelko, J., Ervin, J. and Gruebele, M.: The Cold Denatured Ensemble of Apomyoglobin: Implications for the Early Steps of Folding, *J. Phys. Chem. B* **102** (1998), 1806–1819.
21. Nishii, I., Kataoka, M., Tokunaga, F. and Goto, Y.: Cold Denaturation of the Molten Globule States of Apomyoglobin and a Profile for Protein Folding, *Biochemistry* **33** (1994), 4903–4909.
22. Khorasanizadeh, S., Peters, I. and Roder, H.: Evidence for a Three-State Model of Protein Folding from Kinetic Analysis of Ubiquitin Variants with Altered Core Residues, *Nat. Struct. Biol.* **3** (1996), 193–205.
23. Kollman, P.: Free Energy Calculations: Applications to Chemical and Biochemical Phenomena, *Chem. Rev.* **93** (1993), 2395–2417.
24. Hendrix, D.A., Jarzynski, C.: A Fast Growth, Method of Computing Free Energy Differences, *J. Chem. Phys.* **114** (2001), 5974–5982.
25. Strajbl, M., Sham, Y.Y., Villa, J., Chu, Z.T. and Warshel, A.: Calculations of Activation Entropies of Chemical Reactions in Solution, *J. Phys. Chem. B* **104** (2000), 4578–4584.
26. Pace, C.N.: The Stability of Globular Proteins, *Crit. Rev. Biochem.* **3** (1975), 1–43.
27. Khorasanizadeh, S., Peters, I., Butt, T. and Roder, H.: Folding and Stability of a Tryptophan-Containing Mutant of Ubiquitin, *Biochemistry* **32** (1993), 7054–7063.

28. Qin, J., Ervin, J., Gruebele, M. and Kihara, H.: *Proc. Natl. Acad. Sci.*, USA, submitted for publication (2001).
29. Lazar, G.A., Desjarlais, J.R. and Handel, T.M.: De Novo Design Of the Hydrophobic Core Of Ubiquitin, *Protein Sci.* **6** (1997), 1167–1178.
30. Briggs, M. and Roder, H.: Early Hydrogen Bonding Events in the Folding Reaction of Ubiquitin, *Proc. Natl. Acad. Sci. USA* **89** (1992), 2017–2021.
31. Kramers, H.A.: Brownian Motion In A Field of Force and the Diffusion Model of Chemical Reactions, *Physica* **7** (1940), 284.
32. Weast CRC: *Handbook of Chemistry and Physics*, CRC Press. INC.: Cleveland, OH, 1997.
33. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. and Kiefhaber, T.: The Speed Limit of Protein Folding Measure by Triplet-Triplet Energy Transfer, *Proc. Natl. Acad. Sci. USA* **96** (1999), 9597–9601.
34. Lapidus, L.J., Eaton, W.A. and Hofrichter, J.: Measuring the Rate of Intramolecular Contact Formation in Polypeptides, *Proc. Natl. Acad. Sci. USA* **97** (2000), 7220–7225.