Characterizing the Protein Folding Transition State Using ψ Analysis

Tobin R. Sosnick,*[,†,‡] Bryan A. Krantz,[†,§] Robin S. Dothager,[†,∥] and Michael Baxa[‡,⊥]

Departments of Biochemistry and Molecular Biology and Physics, Institute for Biophysical Dynamics, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637

Received March 17, 2005

Contents

1. Introduction

One of the central problems in structural biology is how the amino acid sequence of a protein codes for its 3D structure. This area of research has become particularly important with the advance of numerous genome projects. Their success provides motivation for "structural genomics" projects, the identification of the structure and function of the entire protein complement of an organism.

Attempts to determine structure from sequence alone have yet to succeed satisfactorily.[1] Elucidation of the mechanism of protein folding can play a major role in structure prediction. In addition, the understanding of the folding mechanisms has tremendous implications to human health as well as biological function. The presence of misfolded, kinetically trapped protein conformers has been implicated in a large number of human diseases (e.g., Alzheimer's, BSE/Mad Cow, Huntington's[2−4]), while the ubiquitous presence of folding chaperones[5−7] testifies to the importance of the kinetics of folding in many cellular activities. The existence of partially disordered proteins such as insulin[8] and protease inhibitors,[9] as well as "natively unfolded" proteins,[10] whose folding is coupled to binding[11] highlights that folding can also be a regulatory mechanism. Likewise, an understanding of the basic science of protein folding is critical to the investigation of complex cellular processes, including those that catalyze the reaction in reverse, such as programmed proteolytic degradation by the proteasome[12,13] or the Clp protease,[14,15] protein import into mitochondria,[16] and bacterial pathogenesis involving toxin entry into cells.[17,18]

Considerable progress has been made on both the experimental and theoretical fronts in mechanistic studies. However, many fundamental issues remain including the following. What do folding pathways look like, and is this even a proper description of the reaction? What structures compose the transition-state ensemble (TSE), and are there structurally disjoint members of the set?

Many of these issues remain open even for simple proteins because most fold in a two-state manner, U ↔ N.[19−22] Under these conditions, intermediates do not accumulate. As a result, often the TSE is the only point on the pathway that can be characterized, although native-state hydrogen exchange methods can describe intermediates that form after the rate-limiting step.[23−29] Unfortunately, TSs cannot be trapped and studied by the usual structural methods, and detailed characterization of the ensemble is difficult with existing methods.

Mutational ϕ analysis has been a major method for characterizing the structure of TS's for protein folding[30−32] and other reactions.[33,34] In this approach, the energetic effect of an amino acid substitution on the folding activation energy relative to its effect on equilibrium stability, quantified as $\phi = \Delta\Delta G^{\ddagger}_{f}/\Delta\Delta G_{eq}$, is interpreted as the extent to which a mutated residue is involved in the formation of the TS. Values of zero and one are taken to indicate that the influence of the side chain is either absent or fully present, respectively, in the TS.

However, in protein folding studies, the reliability of the mutational ϕ analysis to accurately identify TS structure has become a subject of much debate.[35−42] Interpretational issues arise because most ϕ values are fractional, generally lying in the range of 0.1−0.5.[38,39,41,43−49] These intermediate values might be due to either partial structure formation in the TS or the presence of multiple TS structures. Furthermore, if multiple, alternative TS's exist, a destabilizing mutation will reduce the fraction of states in the ensemble that involve this residue and thus generate a lower than expected ϕ value.[40,41,46]

In addition, ϕ values, which largely probe side-chain interactions, generally do not identify the interacting partner(s) in the TS. Examples of non-native interactions generating otherwise reasonable ϕ values have been found.[36,50−52] In general, the effects of an amino acid substitution can depend

* To whom correspondence should be addressed. Phone: (773)-834-0657. Fax: 702-0439. E-mail: trsosnic@midway.uchicago.edu.
† Department of Biochemistry and Molecular Biology, University of Chicago.
‡ Institute for Biophysical Dynamics, University of Chicago.
§ Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115.
∥ Present address: Department of Biochemistry 355 Roger Adams Laboratory University of Illinois Champaign-Urbana 600 South Matthews Urbana, IL 61801.
⊥ Department of Physics, University of Chicago.

10.1021/cr040431q CCC: $59.00
Published on Web 04/22/2006

Tobin R. Sosnick was born in 1961 in Woodland, CA. He received his B.A. degree in Physics from The University of California, San Diego, and his Ph.D. degree in Applied Physics studying momentum distributions in superfluid helium in 1989 from Harvard University working under the supervision of Paul Sokol. He was a postdoctoral research fellow with Jill Trewhella and Cliff Unkefer at Los Alamos National Lab (1988–92) and with S. Walter Englander at The University of Pennsylvania (1992–5). After a year as a research assistant professor at UPENN he joined the faculty at the University of Chicago in 1996.

Robin S. Dothager was born in 1979 in Vandalia, IL. He received his B.A. degree in Chemistry under the supervision of Dr. Jianjun Wang from Southern Illinois University at Carbondale in 2001. He then joined the lab of Dr. Tobin Sosnick as a research technician at the University of Chicago, where he began folding studies of small proteins. In 2003 he commenced his Ph.D. studies under the direction of Dr. Paul Hergenrother at the University of Illinois, Champaign–Urbana. His current research includes identification and mechanistic studies of novel small molecules for treatment of malignant melanoma.

Bryan A. Krantz was born in 1974 in Cincinnati, OH. He received his B.A. and B.S. degrees in English and Chemistry at Emory University in 1996 and his Ph.D. degree at the University of Chicago in 2002, studying the kinetic mechanism of protein folding with Tobin Sosnick. He is continuing postdoctoral research at Harvard University with John Collier, where he is engaged in understanding the processes of protein unfolding and translocation through toxin channels.

Michael C. Baxa was born in 1980 in Plano, TX, before being transplanted to Illinois. He received his B.S. degree in Physics and Mathematics from Western Illinois University in 2002. He joined Tobin Sosnick's lab in 2004 as a Ph.D. candidate in Physics and is currently applying ψ analysis to very fast folding proteins.

on an indeterminate combination of local, long-range, native and even non-native interactions, and secondary structure preferences. As a consequence, the translation of fractional energetic changes quantified by the ϕ value into the language of TS structure is difficult.

Motivated by this issues, a number of methods have been developed recently to probe TS structures including ψ analysis,[35,41,47] kinetic amide-isotope effects,[53−57] and the incorporation of non-natural ester backbones.[58] In ψ analysis, the focus of this review, engineered bi-Histidine (biHis) metal-ion binding sites are introduced one at a time at known positions throughout the protein to stabilize secondary and tertiary structures. The addition of increasing concentrations of metal ions stabilizes the interaction between the two known histidine partners in a continuous way. The analysis thereby is able to quantitatively evaluate the shift in the TSE resulting from the metal-induced stabilization of the biHis site. The translation of a measured ψ value to structure formation is straightforward because the proximity of two known positions is probed. In the following sections, we will discuss this new method along with a comparison to standard mutational techniques as well as the implications to our understanding of TS structures, topologies, and heterogeneity therein.

2. ϕ and ψ Analysis

2.1. ψ Analysis and biHis Binding Sites

In ϕ analysis, as with its forebear Brønsted[59] and Leffler[60] analyses, the stabilization of the TS, $\Delta\Delta G^{\ddagger}_{f}$, due to an energetic perturbation is plotted relative to the change in the equilibrium stability, $\Delta\Delta G_{eq}$ (Figure 1). In folding studies, the proportionality between these two quantities is termed the ϕ value, where $\phi = \Delta\Delta G^{\ddagger}_{f}/\Delta\Delta G_{eq}$, rather than the Brønsted slope, β, which is applicable to covalent bond formation.[59] This linear fit parameter, whether it is ϕ or β, is the proportion of equilibrium energy realized in the interaction or bond at the rate-limiting step.

This rudimentary form assumes a constant, linear slope, implying that the perturbation does not alter the TSE. Rigorously multiple measurements are required to prove a

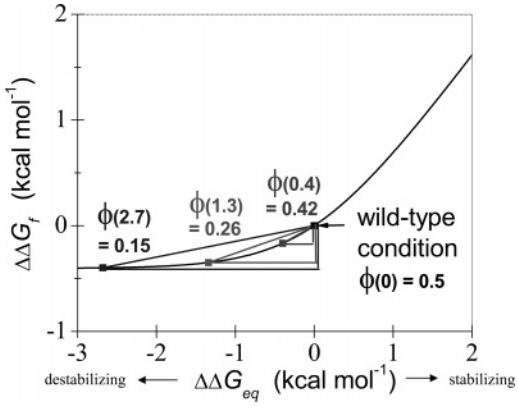

Figure 1. Leffler plot applied to protein folding. In traditional ϕ analysis, the ϕ value is the slope of a two-point fit between the wild-type and mutant protein. When there is TS heterogeneity, ϕ values underestimate the relative importance of an interaction. For a completely nativelike interaction, but which is present in only 50% of the TSE ($K_{eq} = 1$), destabilizing mutations of 0.4, 1.3, and 2.7 kcal mol^{-1} reduce K_{eq} to 0.33, 0.09, and 0.01, respectively. The corresponding ϕ values will be 0.42, 0.26, and 0.15, even though the contribution of the residue to the TS of the wild-type protein is $\phi(0) = 0.5$. Hence, the desirability of having large energetic perturbations to generate accurate ϕ values[37] can be detrimental to correctly assessing the contribution of a residue to the TSE. (Reprinted with permission from ref 35. Copyright 2004 National Academy of Sciences, U.S.A.)

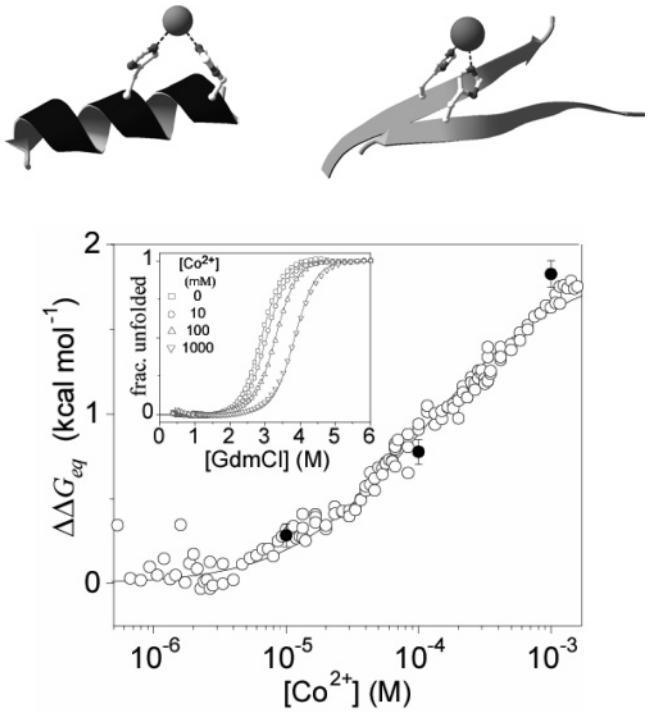

Figure 2. Sample biHis sites located in an i,i+4 arrangement on a helix and across two β strands. Increase in folding stability upon addition of divalent metal ions from folding kinetics (○) and from denaturant melts (●) for the Site *k* Ub variant.[41] (Inset) Standard GdmCl denaturation profiles at different [Co^{2+}] that are fit to a two-state equilibrium model. (Reprinted the permission from ref 41. Copyright 2004 Elsevier.)

linear relationship.[43,48] However, often in folding studies, only a single alanine substitution is examined, which is insufficient to support the linear relationship. Fractional ϕ values, therefore, could be due to TS heterogeneity, fractional energetic interactions, or combinations thereof.

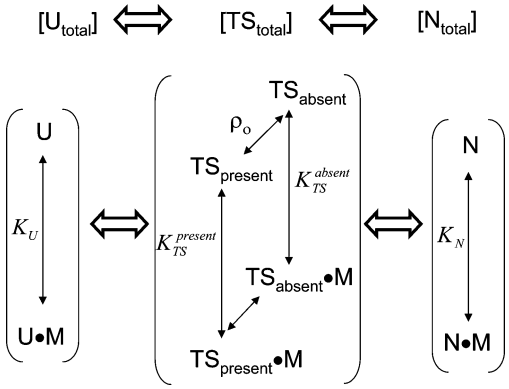

Figure 3. Thermodynamic states considered in ψ analysis. The U, TS_{absent}, $TS_{present}$, and N states bind metal with affinities of K_U, K_{absent}^{TS}, $K_{present}^{TS}$, and K_N, respectively. The degree of TS heterogeneity is defined as $\rho = [TS_{absent}]/[TS_{present}]$.

ψ analysis, which utilizes biHis metal-binding sites, is analogous to mutational ϕ analysis. However, instead of an amino acid substitution, divalent metal-ion concentration is varied to perturb the stability of a specific region (Figure 2). For each biHis variant, a continuous range of values for the change in folding rate and equilibrium stability, $\Delta\Delta G^{\ddagger}_{f}$ and $\Delta\Delta G_{eq}$ are obtained, which can be presented in the Leffler plot.

Our interpretation of such data takes into account the shifts in the native (N), unfolded (U), and TS state populations due to binding of the metal ion to each of these states (Figure 3). In the N state, the biHis site is preformed and binds metal with a dissociation constant K_N. The U ensemble normally does not have the histidines positioned to bind metal. Nevertheless, after a conformational rearrangement with an equilibrium constant γ, the chain can bind metal (M) with a dissociation constant K^*_U

$$U_{absent} + M \xrightarrow{\gamma} U_{present} + M \xleftarrow{K^*_U} U_{present}\cdot M \qquad (1)$$

This reaction is equivalent to the simplified reaction

$$U + M \xleftarrow{K_U} U\cdot M$$

with an effective dissociation constant

$$K_U = K^*_U \frac{\gamma + 1}{\gamma}$$

The increase in protein stability upon addition of metal is given is given by a linked equilibrium expression[61]

$$\Delta\Delta G_{eq} = RT \ln \frac{1 + [M]/K_N}{1 + [M]/K_U} \qquad (2)$$

Folding rates are calculated assuming two classes of TS's, according to whether the biHis site is present ($k^{present}$) or absent (k^{absent}). In the first class, $TS_{present}$, the biHis site is present in a native or near-native geometry with a dissociation constant $K_{TS}^{present}$. The associated backbone structure is folded, for example, in a helical or β-sheet conformation.

In the second class, TS_{absent}, the biHis site is essentially absent. However, just as metal-ion binding can occur in the U state upon a conformational change that brings the two histidines into close proximity, metal-ion binding in the TS_{absent} state can be considered upon a conformational

change, $TS_{absent \rightarrow present}$, with an equilibrium constant λ. Although the $TS_{absent \rightarrow present}$ conformation can bind metal, it is distinct from $TS_{present}$. $TS_{absent \rightarrow present}$ does not have a nativelike backbone structure (otherwise allowing for binding in TS_{absent} would be redundant to the possibility of binding in $TS_{present}$). After accounting for the conformational change, the site has an effective dissociation constant K_{TS}^{absent}. As a result, the model considers two TSs, each having their own effective binding affinities, $K_{TS}^{present}$ and K_{TS}^{absent}.

As per Eyring reaction rate theory,[62] the overall reaction rate is taken to be proportional to the relative populations of the TS and U ensembles, $k_f \propto [TS]/[U]$. The net folding rate is the sum of the rates going down each of the two routes, $k_f = k^{present} + k^{absent}$. Under conditions where metal-binding equilibrium is established faster than the folding rates (which has been observed experimentally via the agreement between $\Delta\Delta G_{eq}$ determined from equilibrium and kinetic measurements[41,47]), the effect of binding can be obtained from the accompanying population shifts of each of the states.

$$k_f = k^{present} + k^{absent} \propto \frac{[TS_{present}] + [TS_{present}\cdot M]}{[U] + [U\cdot M]} + \frac{[TS_{absent}] + [TS_{absent}\cdot M]}{[U] + [U\cdot M]}$$

$$= \frac{[TS_{present}]}{[U]} \frac{1 + [M]/K_{TS}^{present}}{1 + [M]/K_U} + \frac{[TS_{absent}]}{[U]} \frac{1 + [M]/K_{TS}^{absent}}{1 + [M]/K_U} \quad (3)$$

or

$$k_f \equiv k_o^{present} \frac{1 + [M]/K_{TS}^{present}}{1 + [M]/K_U} + k_o^{absent} \frac{1 + [M]/K_{TS}^{absent}}{1 + [M]/K_U} \quad (4)$$

where $k_o^{present} \propto [TS_{present}]/[U]$ and $k_o^{absent} \propto [TS_{absent}]/[U]$ are the rates through each TS class prior to the addition of metal with a ratio $\rho_o = k_o^{absent}/k_o^{present}$. By examining shifts in populations, this treatment avoids any assumptions about possible pathways connecting each of the different metal unbound and bound states.

Normalizing by the rate prior to addition of metal ion ($k_f^o = k_o^{absent} + k_o^{present}$) and substituting for metal dependence of the stability from eq 2 yields

$$\frac{k_f}{k_f^o} = \frac{(k_o^{present} + k_o^{absent}) + (k_o^{present}/K_{TS}^{present} + k_o^{absent}/K_{TS}^{absent})[M]}{(k_o^{present} + k_o^{absent})(1 + [M]/K_U)}$$

$$\equiv 1 - \frac{b}{a} + \frac{b}{a} e^{\Delta\Delta G_{eq}/RT} \quad (5)$$

where

$$a = K_N - K_U$$

$$b = K_N - K_N K_U \left[\frac{k_o^{present} K_{TS}^{absent} + k_o^{absent} K_{TS}^{present}}{K_{TS}^{present} K_{TS}^{absent} (k_o^{present} + k_o^{absent})} \right] \quad (6)$$

The kinetic data, presented in the form of a Leffler plot (Figure 4A), is linear or curved depending upon the relative values of the binding constants and the degree of TS heterogeneity. Despite the apparent complexity of eqs 5 and

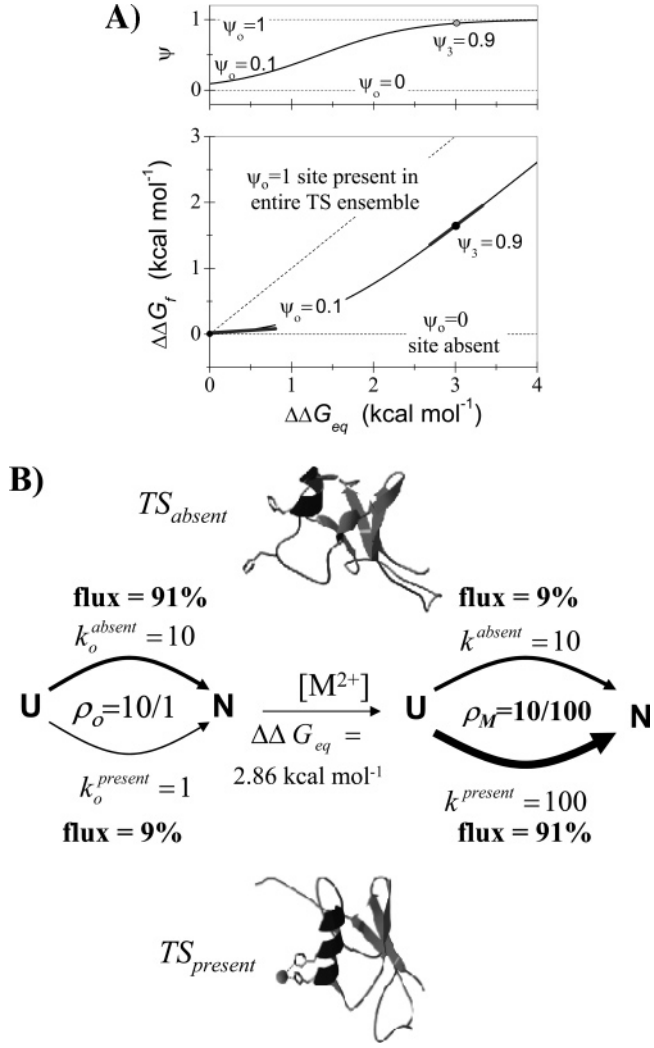

Figure 4. ψ analysis and TS heterogeneity. (A) Generalized Leffler plot for $\psi_o = 0$, 0.1, and 1 (lower panel) and the derivative of each of the traces (upper panel). The $\psi_o = 0.1$ trace is applicable to the scenario shown in B with the thickened lines illustrating the initial condition and after ~3 kcal mol^{-1} of metal-induced stabilization. (B) Application of ψ analysis to a two-route scenario with a helical site with native binding affinity which is formed on 9% of the pathways prior to addition of metal. The absent route contains a TS that has the same binding affinity as the U state. The folding rate for the route with the biHis site present ($k^{present}$, lower pathway) increases from 1 to 100 upon addition of 2.86 kcal mol^{-1} of metal-ion binding energy at 20 °C. This enhancement increases the flux down the metal-ion-stabilized route relative to all other routes (k^{absent}), from $\rho_o = k^{absent}/k_o^{present} = 10/1$, to metal-enhanced condition, $\rho_3 = 10/100$. The corresponding ψ values increase from $\psi_o = 0.1$ to $\psi_3 = 0.9$. The binding energy required to stabilize a TS and switch a minor route to a major route identifies the barrier height for this route relative to that for all other routes. (Reprinted with permission from ref 35. Copyright 2004 National Academy of Sciences, U.S.A.)

6, the data can be fit with a single parameter

$$\Delta\Delta G^{\ddagger}_{f} = RT \ln((1 - \psi_o) + \psi_o e^{\Delta\Delta G_{eq}/RT}) \quad (7)$$

where the slope at the origin is

$$\psi_o = b/a = \frac{K_N}{K_N - K_U}\left(1 - \frac{K_U}{K_{TS}^{absent}}\right) - \frac{K_N K_U}{K_N - K_U}\left(\frac{1}{K_{TS}^{present}} - \frac{1}{K_{TS}^{absent}}\right)\frac{k_o^{present}}{k_o^{present} + k_o^{absent}} \quad (8)$$

The *instantaneous* slope, termed the ψ value, as a function of binding stability, is given by

$$\psi = \frac{\partial\Delta\Delta G^{\ddagger}_{f}}{\partial\Delta\Delta G_{eq}} = \frac{\psi_o}{(1 - \psi_o)e^{-\Delta\Delta G_{eq}/RT} + \psi_o} \quad (9)$$

There are two scenarios where the slope is linear. In the first scenario, the slope is zero across all metal concentration ($\psi = \psi_o = 0$). This behavior occurs when metal-ion binding does not increase the population of the TSE relative to U. The entire TSE lacks the binding site, or more rigorously, the ensemble has the same binding affinity as the unfolded state. At the other limit, the slope is one ($\psi = \psi_o = 1$), indicating that the entire ensemble has the binding site formed in a nativelike manner.

Outside of these two linear situations, the Leffler plot will be curved as added metal continuously increases the population of the TSE. In such cases, the curvature can be due to TS heterogeneity or non-native binding affinity due to a distorted biHis site in a singular TS (D. Goldenberg, private communication) or a combination thereof.

In the heterogeneous scenario, one can consider the simplified situation (Figure 4B), where $TS_{present}$ has the biHis site present with nativelike affinity ($K_{TS}^{present} = K_N$) while TS_{absent} has the site with the unfolded-like affinity ($K_{TS}^{absent} = K_U$). Here only the $TS_{present}$ state is stabilized upon addition of metal ion. The height of the kinetic barrier associated with $TS_{present}$ decreases to the same degree as does the native state, $k^{present} = k_o^{present} e^{\Delta\Delta G_{eq}/RT}$. The instantaneous slope simplifies to the fraction of the TSE which has the biHis site formed at a given metal-ion concentration

$$\psi = \frac{k^{present}}{k^{present} + k_o^{absent}} \quad (10)$$

In the simplified situation, the degree of pathway heterogeneity prior to addition of metal ions is given by ψ_o, the slope at zero stabilization. The Leffler plot exhibits upward curvature as the ψ value increases with added metal binding energy. Generally, ψ values continuously vary between 0 and 1 at the limits of infinite TS destabilization and stabilization, respectively (Figure 4A, upper panel). For example, when the ψ value is 0.5, one-half the TSE has the site formed.

The introduction of the biHis substitution itself alters the stability of the native state by the amount $-\Delta\Delta G_{eq}^{biHis}$. We correct the ψ_o value in order to account for this change in stability by

$$\psi_o^{corr} = \frac{\psi_o}{\psi_o + e^{-\Delta\Delta G_{eq}^{biHis}/RT}(1 - \psi_o)} \quad (11)$$

so that ψ_o^{corr} is the instantaneous Leffler slope at which the metal-ion binding energy is exactly offset by the change in stability due to the biHis substitution. This correction is justified as both metal binding and the biHis substitution affect the same region of the protein. With this correction, the ψ values for all the biHis variants can be combined to construct an accurate representation of the TSE that is appropriate for the wild-type protein prior to any perturbation due to biHis substitution or metal binding.

When one TS class has a distorted biHis site with non-native binding affinity ($K_{TS}^{present} \neq K_N$) while the other class has U-like affinity ($K_{TS}^{absent} = K_U$), the initial slope is the degree of heterogeneity multiplied by an additional factor representing the differential binding affinity between $TS_{present}$ and N

$$\psi_o = \frac{K_N}{K_{TS}^{present}} \frac{K_{TS}^{present} - K_U}{K_N - K_U} \frac{k_o^{present}}{k_o^{present} + k_o^{absent}} \quad (12)$$

Curvature can also occur in a homogeneous scenario when the singular TS has a distorted biHis site with non-native binding affinity. Here the curvature reflects the stabilization of the single TS relative to U. The initial slope is $\psi_o = [K_N(K_{TS} - K_U)]/[K_{TS}(K_N - K_U)]$. In this situation, the interpretation of the two linear behaviors, $\psi = \psi_o = 0$ or 1, remains unchanged, with the biHis site formed with unfolded-like or nativelike manner in the entire TSE, respectively.

When there is both TS heterogeneity and the site is formed in the TS having affinities $K_{TS}^{present} = K_N/\alpha$ and $K_{TS}^{absent} = K_U/\beta$, the initial slope is given by

$$\psi_o = \frac{(1 - \beta)K_N}{K_N - K_U} - \frac{\alpha K_U - \beta K_N}{K_N - K_U} \frac{k_o^{present}}{k_o^{present} + k_o^{absent}} \quad (13)$$

The origin of fractional ψ values, being due to a heterogeneous TSE or a homogeneous TS with a distorted site, may be identified through the use of multiple metal ions that have different coordination geometries. These ions are likely to manifest the same ψ value in the case of TS heterogeneity but different values for a single TS having a distorted biHis site. In the latter case, the different coordination geometries are likely to result in differential fractional binding affinity in the TS, i.e., the metals should stabilize the TS to different extents, relative to the stability they impart to the native state, and thus, they would return different ψ values. In contrast, the ψ value is expected to be the same in the heterogeneous scenario as it only depends on the fraction of the TSE which has nativelike binding affinity and not the magnitude of the binding affinity in the native state which can vary from metal to metal. Support for the heterogeneous scenario was found in the folding of common-type acyl phosphatase. The same fractional ψ value was obtained using Ni^{2+}, Zn^{2+}, and Co^{2+} ions, which prefer octahedral, tetrahedral, and square planar/octahedral coordination geometries,[63] and stabilize the native state by 1.5, 0.9, and 0.7 kcal mol^{-1}, respectively (Pandit, A; Sosnick, T. R. Unpublished data).

An alternative strategy to investigate whether fractional ψ values represent TS heterogeneity is whether the stability of the TS structure with the site present can be altered via

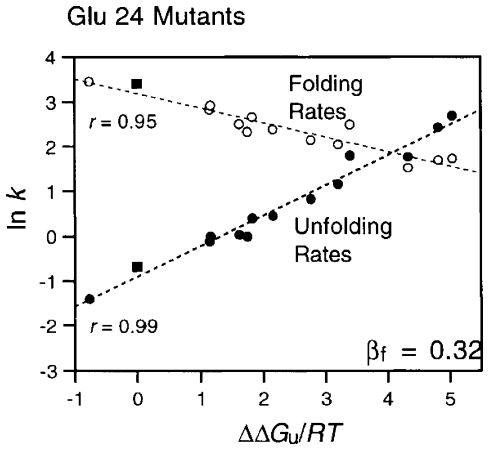

Figure 5. Leffler plot for the Glu24 mutants of Fyn SH3 domain. The fractional, linear behavior indicates the site is partially formed, rather than formed in a subpopulation of the TSE. (Reprinted with permission from ref 48. Copyright 2002 Elsevier.)

mutation (without perturbing the binding site); if the ψ value responds accordingly, as we observed in the coiled coil (see below),[47] the heterogeneity model is the most parsimonious.

In principle, the ψ-analysis formalism can be applied to any perturbation, such as an appropriate series of mutations.[43] In order for this strategy to be successful, each mutation should affect the stability of the TS in a similar manner. Otherwise, the data points cannot be combined and fit using a single ψ value. Finding such mutations may be difficult as most residues are chemically and sterically dissimilar. This dissimilarity may result in ϕ values at a given site being quite disparate when different pairs of residues are compared. Therefore, it is very difficult to obtain a suitable mutational data set with enough high-quality data points to evaluate using the ψ-analysis formalism.

Nevertheless, a constant $\phi^{E24} = 0.32$ was observed over a dozen substitutions at a turn position in Fyn SH3[48] even though these mutations spanned a stability range of 4 kcal mol^{-1} (Figure 5). This linearity indicates a uniform 32% interaction in the TS and not pathway heterogeneity. This result is unexpected as the E24 side chain is involved in electrostatic interactions and hydrogen bonds. It seems unlikely that the same fractional interaction can be maintained across the whole library of chemically distinct substitutions. However, the correlation of $\Delta\Delta G_{eq}$ with the I-site frequencies suggests otherwise, and no reasonable alternative is apparent.

The analysis of metal binding presented here is slightly different than that presented in our earlier papers,[41,47] where curvature was only associated with the heterogeneous model. With the explicit inclusion of the binding affinities in the TS, K_{TS}^{absent} and K_{TS}^{absent}, the f value, $f \equiv \Delta\Delta G_f/\Delta\Delta G_{eq}$, is no longer required. It is generally not constant as it depends on metal-ion concentration, except in the two linear scenarios where $f = 0$ or 1.

Fersht presented an alternative model for his analysis of biHis metal binding data.[64] Although his model results in a similar functional form as the one we have presented here, the curvature in the Leffler plot is interpreted in the context of population shifts accompanying ion binding in the unfolded state. The model omits consideration of new interactions between the metal ion and the TS beyond those which already occurred in the unfolded state, as the microscopic folding rate is assumed to be independent of metal concentration for the route starting from the unfolded state already bound to metal. Furthermore, Fersht proposed models very similar to ours, where the perturbation, be it mutation[43] or ligand binding,[65] is accounted for by altered TS stability, which is not directly considered in his more recent model.

In addition, his more recent proposal that the observed curvature and fractional ψ values are due to unfolded state binding seems inappropriate for two biHis sites located between the amino and carboxy strands in ubiquitin that are separated by 60 residues in the primary sequence (see below). It seems inconsistent to interpret these sites' fractional ψ value being due to metal binding in the denatured state while treating the adjacent site's unity ψ value as binding in a nativelike manner in the TS.

2.2. Utility of biHis Sites

Rather than studying the effects of altered side chains, ψ analysis uses deliberately placed, metal-ion binding sites to probe the fraction of the TSE that has the two histidines positioned to bind metal. The strategy of using engineered metal-ion binding sites in biochemical studies has an extensive history.[66−73] The incorporation of biHis sites on the surface of a protein,[72] as done in ψ analysis, is straightforward compared to the introduction of buried sites with four side-chain ligands inside a protein, which can be an involved design process.[74,75] The three helical I,I+4 biHis sites[76] inserted in the GCN4 coiled coil and ubiquitin (Ub) worked as intended, as did 12 out of the 14 other, nonhelical sites on the surface of Ub.

An important functional distinction exists between metal-ion binding sites located on the surface and those which are buried. With surface sites, the metal-induced stabilization is specific to a particular structural element, such as a helix or hairpin. The region of the protein can be prefolded, with overall stabilization being given by a linked equilibrium, as given in eq 1. With buried sites, however, a metal ion generally is required for the cooperative folding of the entire protein, for example, a zinc finger protein.[69] Accordingly, the binding-induced stabilization is not readily assignable to a particular element. However, shifts in the binding affinity are extremely useful for assaying structural perturbations in other parts of the protein, such as changes in β sheet propensity,[69] according to $\Delta\Delta G^{mutant} = -RT \ln(K_d^{mutant}/K_d^{w\text{-}type})$.

Titrating metal-ion binding biHis sites enables the acquisition of finely spaced, accurate kinetic data required to identify and fit nonlinear Leffler plots. This strategy can be used with endogenous metal binding sites as well.[77] The use of the same protein in ψ analysis enables perfect cancellation of the attempt frequency (i.e., the prefactor in reaction rate theory), as opposed to mutational ϕ-analysis studies which assume the same attempt frequency for the wild-type and mutant proteins ($\Delta\Delta G^{\ddagger}_f = RT \ln [(k_f^{w\text{-}type}/k_f^{mutant})/(k_{attempt}^{w\text{-}type}/k_{attempt}^{mutant})] \approx RT \ln [k_f^{w\text{-}type}/k_f^{mutant}]$). Furthermore, the biHis−metal interaction is unambiguous because the two histidine partners are known. Therefore, the method directly measures a single interaction and reports on chain topology rather than a complicated and indeterminate mixture of local and nonlocal interactions as is frequently the case in mutational studies.

2.3. Application to the GCN4 Coiled Coil

ψ analysis was first applied to the folding of the dimeric and cross-linked versions of the GCN4 α-helical coiled coil

(Figure 6),[47] systems whose TSEs are known to be heterogeneous or homogeneous, respectively.[46] Previously, we deduced that fractional ϕ values in the dimeric version were due to multiple pathways with helix nucleation sites located along the length of the coil.[78] After inserting a destabilizing mutation at one site, which possessed an intermediate ϕ value and was potentially important in the TS, we repeated ϕ analysis at a second site, which previously had a small ϕ value. The new ϕ value at the second site increased significantly, indicating that helix nucleation had shifted toward this region. The heterogeneity was lost when the translational symmetry of the molecule was broken upon introduction of a cross-link at the end of the coiled coil. Nucleation in this unimolecular system occurred at the tethered end.

ψ analysis was conducted on the cross-linked version with the biHis site inserted at amino terminus, while the gly gly cys disulfide tether was placed at either the amino or carboxy termini. The measured ψ values were one and zero for the amino- and carboxy-terminal-linked versions, respectively, indicating that nucleation occurred at whichever end the cross-link was placed. Hence, for the cross-linked coiled coil, folding was homogeneous with pathway selection determined by chain connectivity, in agreement with our mutational studies.[46,78]

The application of ψ analysis to the dimeric protein, where folding was known to be heterogeneous, was a stronger test of the method. At low concentrations of metal, folding rates barely changed and the ψ_o value was about 1/6. Under the assumption that TS binds with either N- or U-like affinity, the ψ_o value represents the ratio of the molecules that nucleated at the biHis region relative to those molecules that nucleated at all other regions. Under this assumption, the biHis route was a very minor pathway prior to the addition of metal. As metal concentration was increased, the amino-terminal region was stabilized. More molecules nucleated at the amino-terminal end, and the ψ value increased to about one-half at high metal concentrations. Although the ψ value should approach unity with further stabilization of the binding site, it did not because of the limited experimental range of metal-ion-induced stabilization ($0-3$ kcal mol^{-1}). A detailed analysis of the metal dependence indicated that the biHis-containing pathway was 80-fold less populated than all other routes in the absence of metal.

Another dimeric version was examined which contained the A24G substitution. The introduction of the destabilizing glycine was anticipated to shift the pathway flux away from this region, so that most nucleation events would encompass the biHis site at the other end of the protein. Indeed, the ψ_o value increased to one-half, indicating that one-half of the nucleation events occurred with the biHis site formed.

A comparison of dimeric A24 and A24G versions indicated that the change in the degree of pathway heterogeneity recapitulated the difference in their equilibrium stability. The A24G mutation was responsible for a shift in amount of flux going through regions other than the amino terminus; the ρ_o value shifted from $\sim$80:1 to $\sim$1:1. The ratio of the heterogeneity in these two versions reflected the loss in stability for this mutation, $\Delta\Delta G_{eq} = RT \ln(\rho_{Ala}/\rho_{Gly})$.

Thus, the A24G mutation was responsible for the $\sim$80-fold change in the ρ_o value, equivalent to 2.5 kcal mol^{-1}. This shift was consistent with the decrease in stability for the mutation in either the biHis or wild-type[46] backgrounds (1.7 ± 0.1 or 2.4 ± 0.1 kcal mol^{-1}, respectively). Hence, ψ analysis quantitatively identified the level of TS heterogeneity determined from multisite Ala/Gly surface mutations.

This result strongly supports the interpretation that the ψ value can report on the degree of TS heterogeneity. The 80-fold increase in the ρ_o value upon substitution at the opposite end of the protein as the biHis site would be expected if the heterogeneous pathway model was applicable with $K_{TS}^{present} = K_N$ and $K_{TS}^{absent} = K_U$. A homogeneous model with nonnative binding affinity in the TS would require that the A24G mutation results in the site acquiring nativelike binding affinity. This would be an unlikely scenario given the 2.5 turns of helix separating the substitution and the biHis site. In addition, this dimeric coiled coil is known to be heterogeneous at the levels reflected by the ψ values.[46] Binding sites introduced into regular helices probably will have nativelike binding affinities in the TS. In which case, fractional ψ values generally will be due to TS heterogeneity.

2.4. Application to Ubiquitin

ψ analysis was next applied to the folding of mammalian Ub,[41] a 76-residue α/β protein which folds in a two-state manner.[20−22,79,80] Fourteen functional biHis sites were introduced at exterior positions along the α helix or across adjacent β strands (Figure 7). Kinetic data were taken at dozens of metal-ion concentrations at a fixed denaturant concentration under either refolding or denaturing conditions. This strategy produces many more points on the Leffler plot compared to the acquisition of entire denaturant-dependent chevron plot at only a few different metal-ion concentrations, as was done in the coiled coil studies (compare Leffler plots in Figure 6 with those in Figure 7).

Of the 14 usable biHis sites, ψ_o values ranged from zero (three sites) to unity (five sites). On the basis of these eight sites where the interpretation of the ψ value was unequivocal, a consensus TS structure was identified (Figures 7−9). The nucleus shared a common, nativelike topology wherein part of the major helix was docked against a β-sheet network composed of different portions of four properly aligned strands.

The remaining six sites had fractional ψ_o values. These sites were located near the edges of the β-sheet network and the amino-terminal portion of the helix. They were formed either in a subfraction of a heterogeneous TSE commensurate with their ψ_o^{corr} value or in the entire ensemble but with a distorted geometry. The coiled coil studies suggested that ψ values for helical sites were likely to represent TS heterogeneity, in this case, the fraying of helix. We believe that the fractional ψ_o values for the β sheet sites also were likely to represent fraying to a substantial degree with an equilibrium constant close to the measured ψ_o value.

A major folding pathway is conjectured incorporating both ψ analysis and other experimental data along with structural considerations (Figure 8). The pathway is largely sequential but contains some degree of heterogeneity. The sites with unity ψ values define a pre-TS structure. The TSE also contains regions with fractional ψ_o values, representing either fractionally formed regions or distorted sites with binding affinity weaker than in the native protein. The next post-TS structure contains these regions fully populated with the biHis sites in a nativelike configuration.

Additional experimental information and structural considerations are used to fill in the missing steps on either side of the TS. With respect to the earliest events, regions of the local B1−B2 hairpin populate nativelike geometries at a low

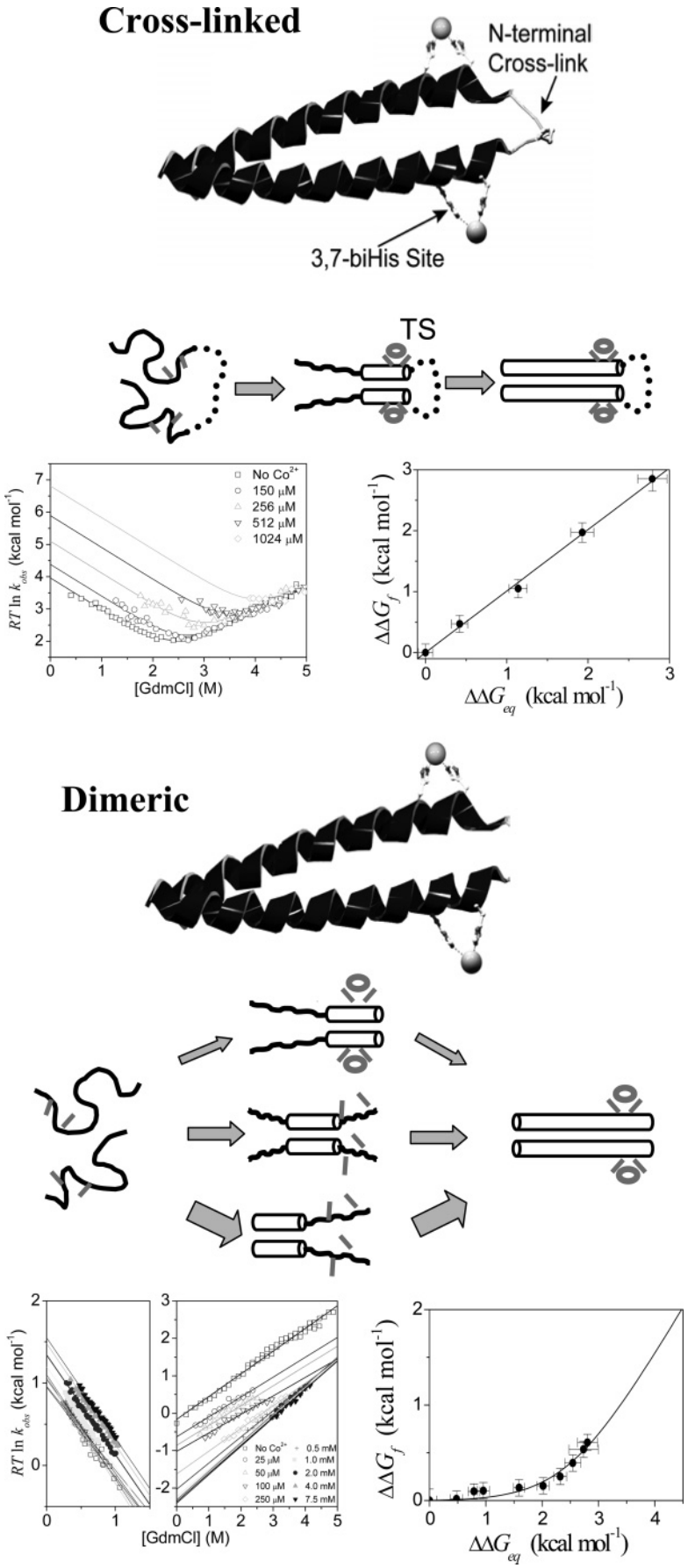

Figure 6. ψ analysis applied to a cross-linked and dimeric GCN4 coiled coil. The cross-linked version folds via a singular TS with nucleation occurring at the tethered end, whereas the dimeric version nucleates a variety of points along the length of the coiled coil. Denaturant dependence of folding kinetics "chevrons" at different metal concentrations and their associated Leffler plots for amino-terminal cross-linked A24G and dimeric A24.

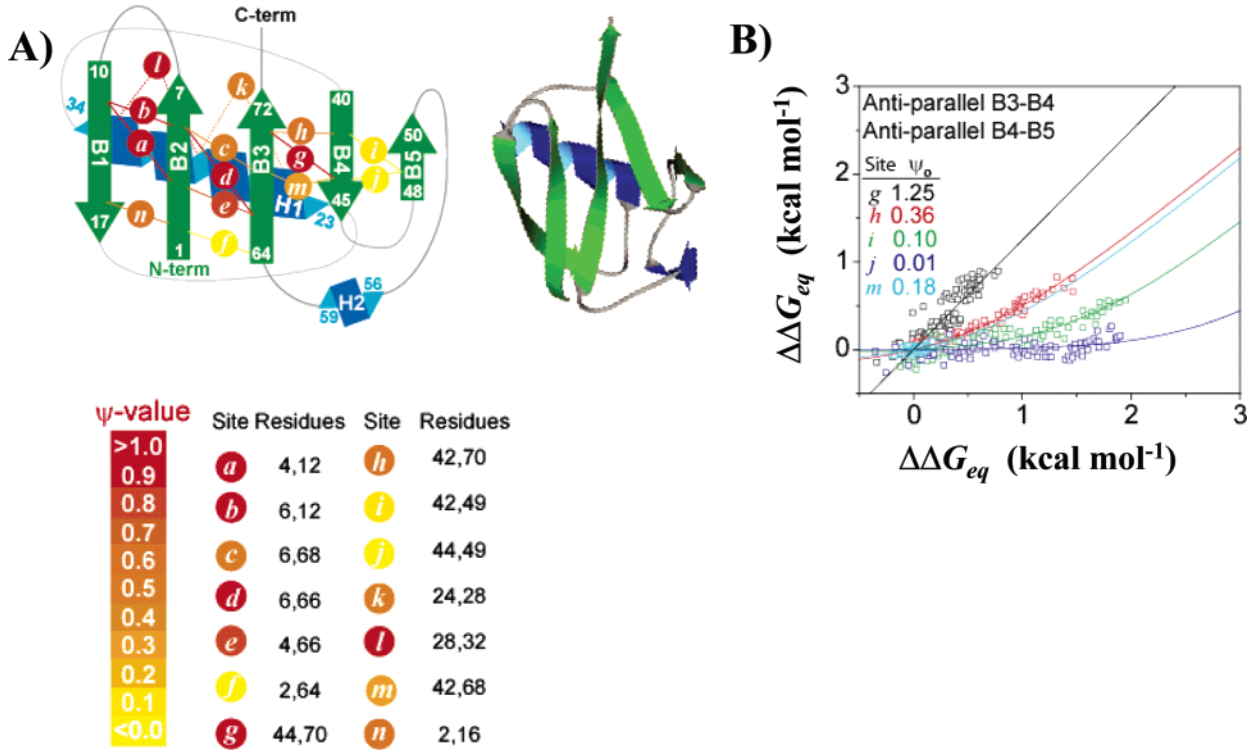

Figure 7. ψ analysis applied to Ub. (A) Schematic representation of biHis sites (circles with italic letters, each site was studied individually) and ψ values. The color intensity represents the value of ψ. Renderings were created in the Swiss-Prot Protein Viewer (Glaxo Wellcome). (B) Leffler plots for five sites illustrating the ψ values ranging from zero to unity. (Reprinted with permission from ref 41. Copyright 2004 Elsevier.)

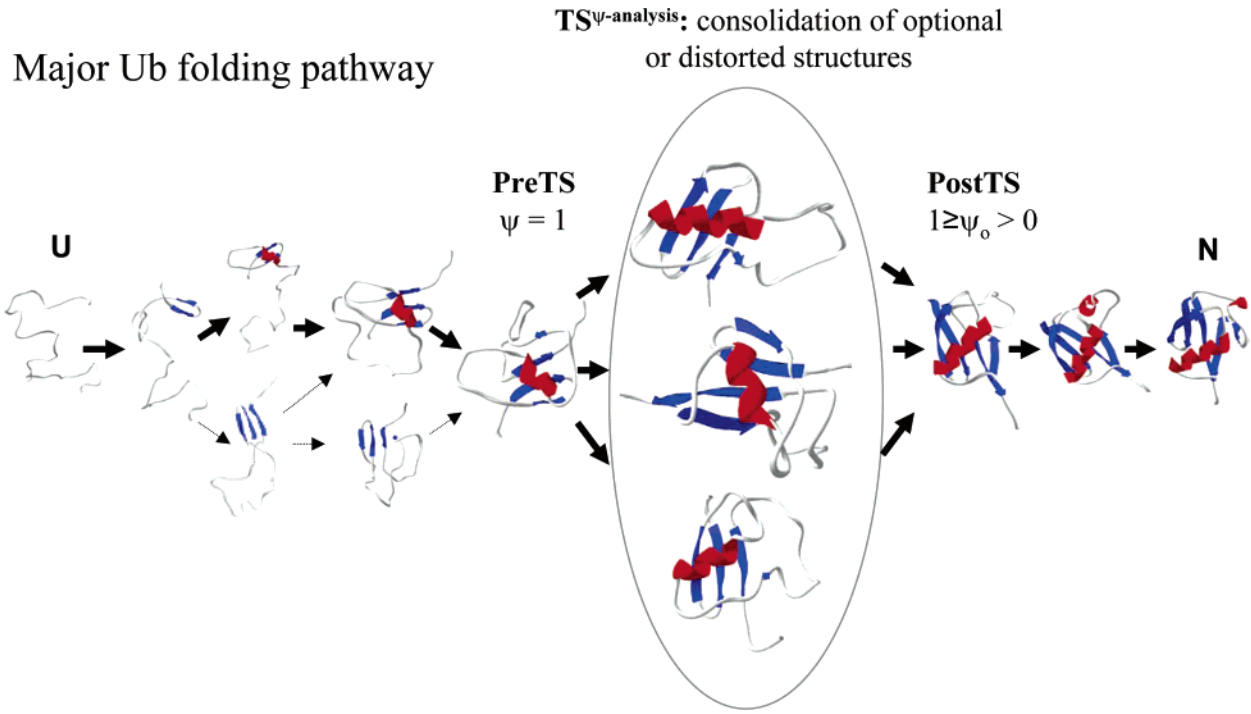

Figure 8. Proposed Ub folding pathway. Ub's folding pathway is described by a heterogeneous TSE that emerges from a conserved nucleus.[41] The upper route on the way to the PreTS structure has more flux than the lower route because of the local nature and increased surface burial of the β hairpin/helix motif compared to the 3-stranded β-sheet structure.

level, ~20%,[81−83] where many of the $\psi_o = 1$ sites are located. In contrast, the helical region has very low intrinsic stability[82] (<3% according to AGADIR[84]). Hence, hairpin formation probably precedes helix formation and is the first major folding event.

Next, the helix docks onto the hairpin to form a local tertiary motif.[82,85] Such a structure is consistent with our kinetic isotope data, which indicate that hydrophobic burial is commensurate with hydrogen-bond formation in the TS.[55] Furthermore, some hairpin NMR resonances change when

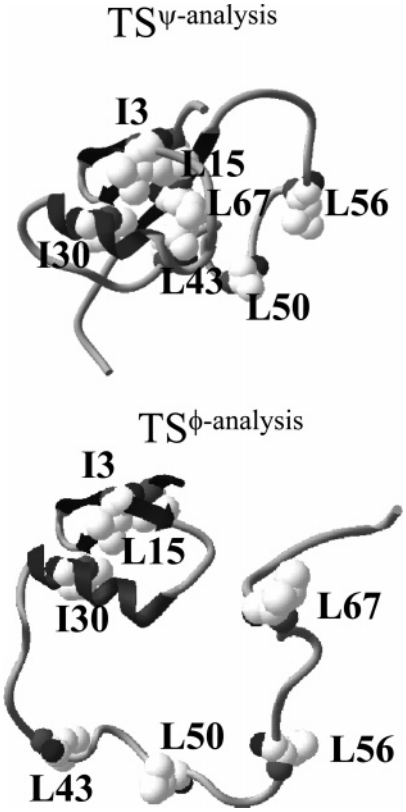

Figure 9. TS identified using ψ and ϕ analysis for Ub. The minimal TS structure determined using ψ analysis is extensive and has the native topology, whereas the smaller and polarized TS determined using ϕ analysis barely defines the ubiquitin fold. (Reprinted with permission from ref 35. Copyright 2004 National Academy of Sciences, U.S.A.)

the fragment is extended to include residues in the helical region.[85] Strand B3 subsequently joins the nascent hairpin−helix nucleus. The final step to the pre-TS structure is the joining of the Strand B4.

Another possible route involves Strand B3 attaching to the B1−B2 hairpin prior to helix formation (lower pathway, dotted arrows in Figure 8). However, the formation of this three-stranded structure would require that the amino and carboxy termini form a parallel β sheet and close a large ~35-residue loop. Furthermore, the sheet would bury less hydrophobic surface than the helix-hairpin nucleus. Hence, this route is less probable than the other route, whereby the helix associates with B1−B2 prior to the joining of B3.

From the post-TS structure identified using ψ analysis, only one turn of the 3_{10} helix and the strand B5 remain unfolded. Hydrogen exchange data on native Ub, which reports on the stability of hydrogen bonds, indicates that two hydrogen bonds, one within the helix and another connecting it to the rest of the protein, are greater than 1.4 kcal mol^{-1} more stable than any of the hydrogen bonds between strand B4 with B5.[86] Hence, the 3_{10} helix likely folds before B5 on the major refolding pathway.

3. Comparison between ϕ and ψ Analysis

We compared the folding TS of Ub identified using ψ analysis to that determined using ϕ analysis.[35] The TSE derived from ψ analysis had an extensive nativelike chain topology, with a four-stranded β-sheet network and a portion of the major helix. According to ϕ analysis, however, the TS was much smaller and polarized[87] with only a local helix/hairpin motif (Figure 9). We found that structured regions can have ϕ values far from unity, the canonical value for such sites, presumably due to structural relaxation of the TS. Consequently, these sites may be incorrectly interpreted as contributing little to the structure of the TS. These results stress the need for caution when interpreting and drawing conclusions from ϕ analysis alone.

Given that mutational ϕ analysis has been a major method for characterizing the structure of TS for protein folding[30,31] and other reactions,[33,34] the origins of the discrepancy between the two methods merit discussion. Interpretational issues arise because most ϕ values are fractional, generally lying in the range of 0.1−0.5.[38,39,41,43−49] These intermediate values might be due to either partial structure formation in the TS or the presence of multiple TS structures. Further, if multiple, alternative TS's exist, a destabilizing mutation will reduce the contribution of states in the ensemble that involve this residue and thus generate a lower than expected ϕ value (Figure 1). For example, our work with the GCN4 coiled coil protein found low single-site ϕ values,[78] which turned out to be due to alternative nucleation positions rather than to the lack of participation by the mutated position.[46,47]

In our comparison between the two methods with Ub, ϕ values were obtained for a set of substitutions in which isoleucine or leucine core residues were changed to alanine on each of the seven major structural elements. These and the ϕ values obtained by Went and Jackson,[87] along with those obtained for surface residues,[41] indicated a small, polarized TS, whereas the ψ-analysis study indicated a much more extensive TS. The most drastic discrepancy was for the L67A substitution on strand β3. The side chain of L67 was estimated to be 74% buried in the TS according to ψ analysis. The four biHis sites located on either side of L67 had ψ values ranging from 0.3 to 1. Nevertheless, the ϕ value for L67A was 0.02 ± 0.01.

The low ϕ values for residues buried in TS probably indicated that the tertiary contacts were not rigid, even when backbone positions were established. For the Leu/Ile → Ala substitutions, the flexible TS may have relaxed to accommodate the void due to the loss of methyl groups in a way that was not available to the more constrained native state. As a result of the TS relaxation, the energetic penalty of the alanine substitution largely would have been ameliorated at the rate-limiting step. Hence, the folding rate was unchanged, even though the substitution was highly destabilizing in the native state.

The ability of the TS to relax to accommodate mutations was supported by other observations. A Ub mutant with seven simultaneous core mutations lost only 1.2 kcal mol^{-1} of stability.[88] Mutations in apomyoglobin regularly gave smaller changes in ΔG_{eq} for the pH 4 intermediate than for native apoMb.[89] Small energetic effects were found for mutations in highly structured regions of BPTI equilibrium intermediates.[38] These and our results suggest that the magnitude of ϕ values can greatly underestimate the native-like character of a residue in the TS.

Another source of uncertainty arises when ϕ values, which reflect energetic perturbations, are used to describe the structure of the TS. Only in a few cases does a clear correspondence between the ϕ value and structure formation exist, for example, in Ala to Gly mutations on the surface of a helix[41,44,46,78,90−92] or in exposed turns that form specific interactions.[93] In other circumstances, converting energetic changes into structure is difficult. Nevertheless, a recent study

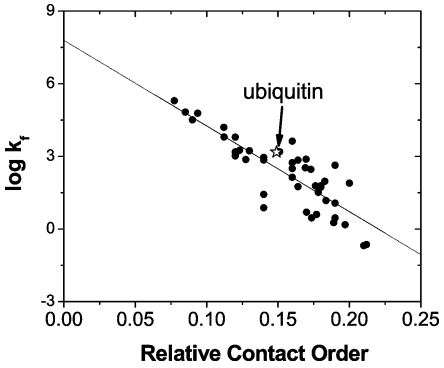

Figure 10. Correlation between relative contact order and folding rates. Each data point represents an individual two-state protein whose folding properties were measured under comparable solution conditions as well as the RCO of the native structure. The line indicates the best linear fit to the data, which we believe represents proteins which form ~80% of the native topology at the TS. Data obtained from ref 140.

of the three-strand PIN WW domain probing hydrogen-bond formation (site-resolved amide to ester changes) did observe qualitative agreement between mutational ϕ values.[58]

The interpretation of ϕ values is further complicated by the possibility of non-native interactions.[50,51] Bai and coworkers elegantly trapped and solved the structure of a folding intermediate for apocytochrome b_{562} and found that the core repacked with a number of non-native interactions.[52] Mutations at these sites produced thermodynamic ϕ values ranging from 0 to 1.[36] Hence, non-native interactions can produce otherwise normal ϕ values which are likely to be misinterpreted.

As a result of fractional interactions, structural relaxation of the TS, TS heterogeneity, and non-native interactions, most ϕ values will cluster in the range 0.1−0.5[38,39,43−48] rather than taking on the canonical values of zero and one. In particular, a highly structured region can have ϕ values far from unity. These values can even be smaller than values resulting from non-native interactions. Furthermore, inferring structure from ϕ values, which reflect energetic effects, is difficult. As a result, the use of ϕ values to identify a TS structure, much less the degree of TS heterogeneity, is problematic. Consequently, the frequent classification of a TS in other proteins as polarized, diffuse, or expanded based upon ϕ values alone needs to be reevaluated. See ref 42 for further discussion of these matters.

4. Topology and the Transition State

We proposed that the folding of small proteins is a nucleation process where the critical element in the TS is the formation of a coarse version of the native chain topology or overall fold.[94−96] Topology is established by pinning the chain with apolar side-chain interactions. This proposal seems to be validated by the success of the correlation between relative contact order (RCO), which reflects the ratio between local and nonlocal contacts, and log k_f (Figure 10) subsequently found by Plaxco et al.[97] This correlation and similar correlations described by others[98−103] points to a two-state folding reaction being limited by an initial conformational search to find the nativelike topology. This powerful observation argues for the importance of topology and that a commonality exists between the folding mechanism of different structural classes.[98−100,104−107]

ψ analysis identifies residue−residue contacts, which makes it particularly well suited to addressing the origin of the correlation between k_f and RCO, reflecting properties of the TS and the native state, respectively. Recapitulating the trend observed for other small proteins,[97] Ub's folding rate correlates with the degree to which the overall fold is

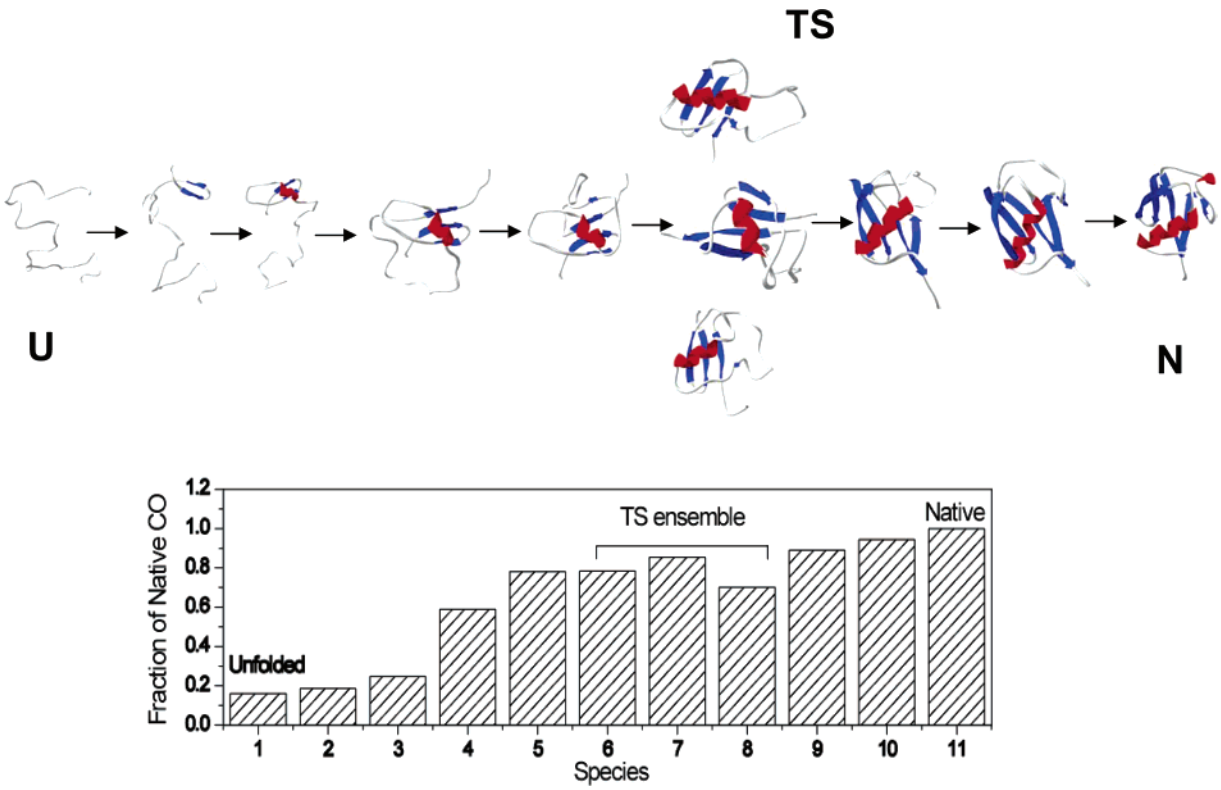

Figure 11. Relative contact order on the proposed Ub folding pathway. Values for each species are normalized to the RCO of the native protein. (Reprinted with permission from ref 41. Copyright 2004 Elsevier.)

achieved, as defined by the average sequence distance between long-range contacts or RCO. The RCO of Ub's TSE is about 80% of the RCO of the native protein (Figure 11). The major increase in RCO occurs with docking of strands B1 and B3, which are the amino and carboxy termini.

The similarity of the gross topology of the TS and the native state supports the observed correlation between the folding rate constant and the RCO of the native state. Because k_f directly relates to the stability of the TS ($k_f \propto e^{-\Delta G^{\ddagger}_f/RT}$) and not the native state, k_f should only reflect properties of the TS. The present demonstration that the TS of Ub has largely adopted the native fold provides a connection between folding rates and the topology of the native state.

The TSE identified using ψ analysis in combination with the RCO trend suggests an intriguing proposition: For proteins which obey the known RCO correlation, their TS's will have ~80% of the native RCO. The rationale for this proposition is as follows. The empirical contact order correlation is $\log k_f = 8.3 - 39\mathrm{RCO}^{\mathrm{native}}$ (Figure 10). For this correlation to be valid, the topology of the TS must closely resemble that of the native state. Consistently, Ub's TS has a very nativelike topology, with a $\mathrm{RCO}^{\mathrm{TS}} \approx 0.8\mathrm{RCO}^{\mathrm{native}}$ (Figure 11), although it is slightly above the best fit line. If the RCO correlation is to hold for a variety of proteins, their TS's likewise should have $\mathrm{RCO}^{\mathrm{TS}} \approx 0.8\mathrm{RCO}^{\mathrm{native}}$ in order for them to be on the same correlation line. That is, Ub's value calibrates the connection between $\mathrm{RCO}^{\mathrm{TS}}$ and RCO of the native state.

Another rationale is illustrated with a counter-example. If a protein only needs to form part of the native topology (e.g., $\mathrm{RCO}^{\mathrm{TS}} \approx 0.5\mathrm{RCO}^{\mathrm{native}}$), it would fold faster than predicted from the observed trend because forming this simplified TS is less costly than the typical ones obeying $\mathrm{RCO}^{\mathrm{TS}} \approx 0.8\mathrm{RCO}^{\mathrm{native}}$.

The high value for $\mathrm{RCO}^{\mathrm{TS}}$ also restricts the degree to which a TS can be small and polarized.[108−116] From the ϕ values for such proteins, we estimate $\mathrm{RCO}^{\mathrm{TS}}$ is often below $0.5\mathrm{RCO}^{\mathrm{native}}$, although the precise number depends on the threshold that a ϕ value is considered to be a contact. At face value, these results would seem to be incompatible with the proposition that $\mathrm{RCO}^{\mathrm{TS}} \approx 0.8\mathrm{RCO}^{\mathrm{native}}$.

However, some caveats exist in the identification of a small, polarized TS based solely upon medium to high ϕ values on one side of the protein. First, ϕ analysis leads to an assignment of a small, polarized TS in ubiquitin, whereas the TS defined by the unequivocal $\psi = 1$ sites is much more extensive.[35] As Schmid et al. astutely noted in their studies "the transition state of CspB folding is polarized energetically, but it does not imply that one part of the protein is folded and the other one is unfolded. Rather, it means that the positions that have reached a native-like energetic environment in the transition state are distributed unevenly."[108] That is, *energetically* polarized does not necessarily mean *structurally* polarized.

Second, the low level of structure formation inferred from ϕ analysis in these situations seems inconsistent with the high percentage of surface burial in the TS. The third caveat is that many high ϕ values in polarized TSs are associated with turns.[93,108−110,114] These high values may not yield a complete picture of the topology of the TS. For example, Serrano et al. concluded that the three SH3 homologues, SSo7D, src- and α-spectrin, fold via different TSs as their turns have different ϕ values.[105] However, an alternative interpretation is that the overall TS topology is similar in all three proteins but the turns are only folded to the degree required for the chain to turn around. Not all turns have to be nativelike in order for the chain to double-back on itself. If true, the sensitivity of the ϕ values is more a reflection of the specifics of the turn rather than the topology of TS. For example, the distal β-hairpin in src-SH3 with high ϕ values is a tight turn,[109] which is quite sensitive to mutation. The corresponding turn in SSo7D contains three flexible glycines, which are not as sensitive to mutation, and this turn has low ϕ values.[105] Hence, ϕ values could be different for this turn despite the two proteins having similar TS topologies.

5. Transition-State Heterogeneity

An issue which has sparked much debate is whether multiple folding pathways exist.[117,118] The traditional biochemist views folding as a determinate A-to-B-to-C process with a specific series of events. Theoretical work has led to a funnel-like picture in which folding occurs via structurally distinct, parallel routes.[101,111,118−127] Despite much appeal, only limited evidence supports the possibility of pathway heterogeneity in the absence of misfolding or proline isomerization issues. Exceptions for small proteins include the dimeric GCN4 coiled coil,[46,47] titin I27,[128] barstar,[129] WW domain,[130] CspB,[131] and Ub.[41]

Our Ub studies highlighted the importance of defining what is meant by "multiple pathways" and "TS heterogeneity" (Figure 12). We found that Ub folds through a nativelike TSE with a common nucleus. However, heterogeneity probably exists wherein peripheral regions are differentially populated according to their relative stability (Figure 12, center). At a coarse level, a single TS exists. However, there are likely different structures within the TSE. Hence, the Ub results can support either the single-pathway or the multiple-pathway view, depending upon the level one wishes to define "multiple pathways". Therefore, it is critical to distinguish between one scenario where there is "microscopic heterogeneity" around a consensus core and another where there are *structurally disjoint* members of the TSE (Figure 12, right).

The dimeric GCN4 coiled coil has disjoint nuclei. For titan, heterogeneity was identified by an upturn in the unfolding arm of the chevron plot.[128] The pattern of ϕ values indicates that the larger TSE which predominates at low denaturant concentration contains an extra β strand and some additional structure formed around the periphery of the central nucleus. The type of heterogeneity seems closer to microscopic heterogeneity described earlier rather than an example of disjoint nuclei.

Other research investigating multiple pathways includes the Serrano et al.[45,132] and Baker et al.[133] studies of SH3 domains. They found no evidence for a shift in pathway upon destabilization of elements formed in the TS. Similarly, loop insertion studies did not detect pathway heterogeneity.[133,134] Although homologues of SH3 fold through different TS's,[105,135] this does not mandate that there are multiple TS's for a single protein sequence. Topological changes sometimes[113,136−138] but not always[133,139] result in different TS's. Baker et al.[135] observed that the topologically identical Protein G and Protein L nucleate at either of two different, but symmetrically related, turns. Mutational studies indicate that nucleation can occur at either turn and reflects their relative stability. Symmetric proteins, such as Protein G and Protein L, probably can fold via TS's which are symmetrically

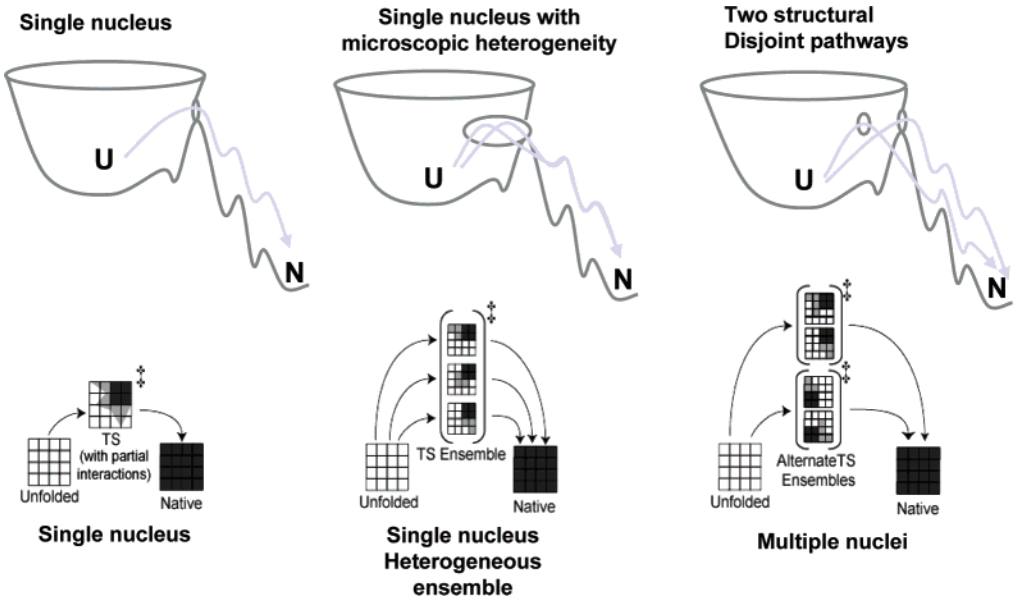

Figure 12. Classes of transition-state heterogeneity. Folding may occur via a singular essential TS nucleus with some partially formed interactions (left). Here, some residues are completely unfolded (white boxes) or absolutely required in a given nucleus (dark gray boxes), while others may have a fractional side-chain interaction in the TS (light gray boxes). Folding may also occur through a structurally heterogeneous ensemble where some residues are critical for the folding nucleus (dark gray boxes) but different groups of structures may also exist at the TS (middle). Here, the TS can exhibit "microscopic heterogeneity" (light gray boxes), e.g., helical fraying, but a conserved folding nucleus. Alternatively, the nuclei can be structurally disjoint (right), each with a diverse set of necessary structures comprising distinct nuclei (dark gray boxes).

related (e.g., mirror images).[118,135] Similarly, we found multiple nuclei for the dimeric coiled coil, which has translational symmetry. However, there is a single pathway when the symmetry is broken upon insertion of a cross-link at one end of the coiled coil.[46]

6. Summary

Protein folding has been both blessed and cursed by the fact that many proteins do not have populated folding intermediates. The lack of stable intermediates simplifies the analysis as the reaction is kinetically two state. However, the lack of intermediates that can be trapped and characterized largely precludes elucidation of the individual folding steps. Therefore, research efforts have largely focused on the rate-limiting step and TSE. This high-energy ensemble can be structurally characterized by methods that seek to quantify the relative energetic contributions of individual side chains in the TS as compared to the native state. The degree of energetic "nativeness" is correlated to structure. Principally, this type of analysis is achieved through alanine-scanning mutagenesis coupled to ϕ analysis.

However, new methods are required to better describe the structure of the TS because alanine-scanning mutagenesis relies on a limited alphabet of amino acids, which are too sterically and chemically dissimilar to reliably indicate the energetic contribution of interactions in the TS. Furthermore, the numerous interacting partners in a given tertiary interaction often cannot be disentangled, encumbering the interpretation of a particular ϕ value.

We have demonstrated that biHis sites incorporated in both α helices and β sheets may be stabilized by the addition of divalent metal ions so that detailed study of the energy of one specific interaction in the TS may be achieved. We termed this kinetic treatment ψ analysis. Rigorous mathematical treatment of the metal-ion binding has indicated it is challenging to divorce pathway heterogeneity from non-native metal-ion affinity in the TS. However, ψ analysis of the dimeric GCN4 coiled coil indicated that the pathway was heterogeneous and confirmed a previous independent study that employed ala-gly mutagenesis. Hence, the biHis methodology can accurately quantify TS heterogeneity. For Ub, we found that the TS is much more structured than expected when ala-scanning ϕ analysis was applied. The large amount of structure and nativelike topology in Ub's TS partly rationalizes the correlation between folding rate and relative contact order found for two-state folding proteins.

A synthesis of ψ-analysis methods and the ability to incorporate non-natural amino acids that are capable of binding metal ions with heightened affinities will increase the energetic space that can be probed. From these and future ψ-analysis studies, an extremely detailed picture of the TSE and pathways can be obtained. As a result, more rigorous comparisons between experiment and theoretical simulations can be conducted to the benefit of both areas.

7. Abbreviations

biHis	bi-histidine divalent metal-ion binding site
TS	transition state
Ub	mammalian ubiquitin
ψ_o	initial slope in the Leffler plot
ψ_o^{corr}	ψ_o value corrected for the change in the protein stability due to the bi-His substitution
ρ_o	the ratio of rates in the absence of metal ions for pathways in which the bi-histidine site is present to those in which bi-histidine site is absent

8. Acknowledgments

We thank S. W. Englander, D. Goldenberg, A. R. Fersht, Y. Bai, N. Kallenbach, R. L. Baldwin, L. Mayne, V. Pande, K. Plaxco, S. Jackson, T. Pan, and members of our group for extensive discussions. This work was supported by grants from the NIH and The Packard Foundation Interdisciplinary

Science Program (T.R.S., P. Thiyagarajan, S. Berry, D. Lynn, S. Meredith).

9. References

(1) Venclovas, C.; Zemla, A.; Fidelis, K.; Moult, J. *Proteins* **2003**, *53* (Suppl 6), 585.
(2) Koo, E. H.; Lansbury, P. T., Jr.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9989.
(3) Prusiner, S. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13363.
(4) Kelly, J. W. *Curr. Opin. Struct. Biol.* **1998**, *8*, 101.
(5) Weissman, J. S.; Kashi, Y.; Fenton, W. A.; Horwich, A. L. *Cell* **1994**, *78*, 693.
(6) Hartl, F. U.; Hlodan, R.; Langer, T. *Trends Biochem. Sci.* **1994**, *19*, 20.
(7) Todd, M. J.; Lorimer, G. H.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4030.
(8) Hua, Q. X.; Ladbury, J. E.; Weiss, M. A. *Biochemistry* **1993**, *32*, 1433.
(9) Cunningham, E. L.; Jaswal, S. S.; Sohl, J. L.; Agard, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11008.
(10) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573.
(11) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54.
(12) Braun, B. C.; Glickman, M.; Kraft, R.; Dahlmann, B.; Kloetzel, P. M.; Finley, D.; Schmidt, M. *Nat. Cell Biol.* **1999**, *1*, 221.
(13) Strickland, E.; Hakala, K.; Thomas, P. J.; DeMartino, G. N. *J. Biol. Chem.* **2000**, *275*, 5565.
(14) Hoskins, J. R.; Sharma, S.; Sathyanarayana, B. K.; Wickner, S. *Adv. Protein Chem.* **2001**, *59*, 413.
(15) Kenniston, J. A.; Baker, T. A.; Fernandez, J. M.; Sauer, R. T. *Cell* **2003**, *114*, 511.
(16) Wiedemann, N.; Frazier, A. E.; Pfanner, N. *J. Biol. Chem.* **2004**, *279*, 14473.
(17) van der Goot, F. G.; Lakey, J.; Pattus, F.; Kay, C. M.; Sorokine, O.; Van Dorsselaer, A.; Buckley, J. T. *Biochemistry* **1992**, *31*, 8566.
(18) Krantz, B. A.; Trivedi, A. D.; Cunningham, K.; Christensen, K. A.; Collier, R. J. *J. Mol. Biol.* **2004**, *344*, 739.
(19) Jackson, S. E. *Fold. Des.* **1998**, *3*, R81.
(20) Krantz, B. A.; Sosnick, T. R. *Biochemistry* **2000**, *39*, 11696.
(21) Krantz, B. A.; Mayne, L.; Rumbley, J.; Englander, S. W.; Sosnick, T. R. *J. Mol. Biol.* **2002**, *324*, 359.
(22) Jacob, J.; Krantz, B.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *338*, 369.
(23) Vu, N. D.; Feng, H.; Bai, Y. *Biochemistry* **2004**, *43*, 3346.
(24) Takei, J.; Pei, W.; Vu, D.; Bai, Y. *Biochemistry* **2002**, *41*, 12308.
(25) Chu, R.; Pei, W.; Takei, J.; Bai, Y. *Biochemistry* **2002**, *41*, 7998.
(26) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192.
(27) Maity, M.; Maity, H.; Englander, S. W. *Biophys. J.* **2004**, *86*, 498a.
(28) Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213.
(29) Xu, Y.; Mayne, L.; Englander, S. W. *Nat. Struct. Biol.* **1998**, *5*, 774.
(30) Matthews, C. R. *Methods Enzymol.* **1987**, *154*, 498.
(31) Fersht, A. R.; Matouschek, A.; Serrano, L. *J. Mol. Biol.* **1992**, *224*, 771.
(32) Goldenberg, D. P. In *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman: New York, 1992.
(33) Fersht, A. R.; Leatherbarrow, R. J.; Wells, T. N. C. *Nature* **1986**, *322*, 284.
(34) Leatherbarrow, R. J.; Fersht, A. R.; Winter, G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7840.
(35) Sosnick, T. R.; Dothager, R. S.; Krantz, B. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17377.
(36) Feng, H.; Vu, N. D.; Zhou, Z.; Bai, Y. *Biochemistry* **2004**, *43*, 14325.
(37) Sanchez, I. E.; Kiefhaber, T. *J. Mol. Biol.* **2003**, *334*, 1077.
(38) Bulaj, G.; Goldenberg, D. P. *Nat. Struct. Biol.* **2001**, *8*, 326.
(39) Ozkan, S. B.; Bahar, I.; Dill, K. A. *Nat. Struct. Biol.* **2001**, *8*, 765.
(40) Fersht, A. R.; Sato, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7976.
(41) Krantz, B. A.; Dothager, R. S.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *337*, 463.
(42) Raleigh, D. P.; Plaxco, K. W. *Protein Pept. Lett.* **2005**, *12*, 117.
(43) Fersht, A. R.; Itzhaki, L. S.; elMasry, N. F.; Matthews, J. M.; Otzen, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10426.
(44) Kim, D. E.; Yi, Q.; Gladwin, S. T.; Goldberg, J. M.; Baker, D. *J. Mol. Biol.* **1998**, *284*, 807.
(45) Martinez, J. C.; Pisabarro, M. T.; Serrano, L. *Nat. Struct. Biol.* **1998**, *5*, 721.
(46) Moran, L. B.; Schneider, J. P.; Kentsis, A.; Reddy, G. A.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10699.
(47) Krantz, B. A.; Sosnick, T. R. *Nat. Struct. Biol.* **2001**, *8*, 1042.
(48) Northey, J. G.; Maxwell, K. L.; Davidson, A. R. *J. Mol. Biol.* **2002**, *320*, 389.
(49) Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Nat. Struct. Biol.* **1996**, *3*, 193.
(50) Cho, J. H.; Sato, S.; Raleigh, D. P. *J. Mol. Biol.* **2004**, *338*, 827.
(51) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *Nat. Struct. Biol.* **2002**.
(52) Feng, H.; Takei, J.; Lipsitz, R.; Tjandra, N.; Bai, Y. *Biochemistry* **2003**, *42*, 12461.
(53) Kentsis, A.; Sosnick, T. R. *Biochemistry* **1998**, ***37***, 14613.
(54) Parker, M. J.; Clarke, A. R. *Biochemistry* **1997**, *36*, 5786.
(55) Krantz, B. A.; Srivastava, A. K.; Nauli, S.; Baker, D.; Sauer, R. T.; Sosnick, T. R. *Nat. Struct. Biol.* **2002**, *9*, 458.
(56) Krantz, B. A.; Moran, L. B.; Kentsis, A.; Sosnick, T. R. *Nat. Struct. Biol.* **2000**, *7*, 62.
(57) Meisner, W. K.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13478.
(58) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101.
(59) Brønsted, J. N.; Pedersen, K. *Z. Phys. Chem.* **1924**, *A108*, 185.
(60) Leffler, J. E. *Science* **1953**, *107*, 340.
(61) Sharp, K. A.; Englander, S. W. *Trends Biochem. Sci.* **1994**, *19*, 526.
(62) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107.
(63) Jia, Y. Q. *J. Solid State Chem.* **1991**, *95*, 184.
(64) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17327.
(65) Sancho, J.; Meiering, E. M.; Fersht, A. R. *J. Mol. Biol.* **1991**, *221*, 1007.
(66) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11255.
(67) Liu, H.; Schmidt, J. J.; Bachand, G. D.; Rizk, S. S.; Looger, L. L.; Hellinga, H. W.; Montemagno, C. D. *Nat. Mater.* **2002**, *1*, 173.
(68) Goedken, E. R.; Keck, J. L.; Berger, J. M.; Marqusee, S. *Protein Sci.* **2000**, *9*, 1914.
(69) Kim, C. A.; Berg, J. M. *Nature* **1993**, *362*, 267.
(70) Webster, S. M.; Del Camino, D.; Dekker, J. P.; Yellen, G. *Nature* **2004**, *428*, 864.
(71) Lu, Y.; Berry, S. M.; Pfister, T. D. *Chem. Rev.* **2001**, *101*, 3047.
(72) Higaki, J. N.; Fletterick, R. J.; Craik, C. S. *TIBS* **1992**, *17*, 100.
(73) Morgan, D. M.; Lynn, D. G.; Miller-Auer, H.; Meredith, S. C. *Biochemistry* **2001**, *40*, 14020.
(74) Benson, D. E.; Wisz, M. S.; Hellinga, H. W. *Curr. Opin. Biotechnol.* **1998**, *9*, 370.
(75) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11255.
(76) Ghadiri, M.; Choi, C. *J. Am. Chem. Soc.* **1990**, *112*, 1630.
(77) Kuwajima, K.; Mitani, M.; Sugai, S. *J. Mol. Biol.* **1989**, *206*, 547.
(78) Sosnick, T. R.; Jackson, S.; Wilk, R. M.; Englander, S. W.; DeGrado, W. F. *Proteins* **1996**, *24*, 427.
(79) Went, H. M.; Benitez-Cardoza, C. G.; Jackson, S. E. *FEBS Lett.* **2004**, *567*, 333.
(80) Vallee-Belisle, A.; Turcotte, J. F.; Michnick, S. W. *Biochemistry* **2004**, *43*, 8447.
(81) Zerella, R.; Chen, P. Y.; Evans, P. A.; Raine, A.; Williams, D. H. *Protein Sci.* **2000**, *9*, 2142.
(82) Jourdan, M.; Searle, M. S. *Biochemistry* **2000**, *39*, 12355.
(83) Zerella, R.; Evans, P. A.; Ionides, J. M.; Packman, L. C.; Trotter, B. W.; Mackay, J. P.; Williams, D. H. *Protein Sci.* **1999**, *8*, 1320.
(84) Munoz, V.; Serrano, L. *Biopolymers* **1997**, *41*, 495.
(85) Chen, P. Y.; Gopalacushina, B. G.; Yang, C. C.; Chan, S. I.; Evans, P. A. *Protein Sci.* **2001**, *10*, 2063.
(86) Pan, Y.; Briggs, M. S. *Biochemistry* **1992**, *31*, 11405.
(87) Went, H. M.; Jackson, S. E. *Protein Eng. Des. Sel.* **2005**, *18*, 229.
(88) Benitez-Cardoza, C. G.; Stott, K.; Hirshberg, M.; Went, H. M.; Woolfson, D. N.; Jackson, S. E. *Biochemistry* **2004**, *43*, 5195.
(89) Kay, M. S.; Baldwin, R. L. *Nat. Struct. Biol.* **1996**, *3*, 439.
(90) Sato, S.; Religa, T. L.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6952.
(91) Matthews, J. M.; Fersht, A. R. *Biochemistry* **1995**, *34*, 6805.
(92) Zitzewitz, J. A.; Ibarra-Molero, B.; Fishel, D. R.; Terry, K. L.; Matthews, C. R. *J. Mol. Biol.* **2000**, *296*, 1105.
(93) McCallister, E. L.; Alm, E.; Baker, D. *Nat. Struct. Biol.* **2000**, *7*, 669.
(94) Sosnick, T. R.; Mayne, L.; Hiller, R.; Englander, S. W. Peptide and Protein Folding Workshop, Philadelphia, PA, 1995; p 52.
(95) Sosnick, T. R.; Mayne, L.; Englander, S. W. *Proteins* **1996**, *24*, 413.
(96) Sosnick, T. R.; Englander, S. W. In *Dynamics and the Problem of Recognition in Biological Macromolecules*; Jardetzky, O., Lefevre, J.-F., Eds.; Plenum Press: New York, 1996; Vol. 288.
(97) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985.
(98) Goldenberg, D. P. *Nat. Struct. Biol.* **1999**, *6*, 987.
(99) Bai, Y.; Zhou, H.; Zhou, Y. *Protein Sci.* **2004**, *13*, 1173.
(100) Ivankov, D. N.; Garbuzynskiy, S. O.; Alm, E.; Plaxco, K. W.; Baker, D.; Finkelstein, A. V. *Protein Sci.* **2003**, *12*, 2057.
(101) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Biochemistry* **1994**, *33*, 10026.

(102) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869.
(103) Guo, Z. Y.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.
(104) Alm, E.; Morozov, A. V.; Kortemme, T.; Baker, D. *J. Mol. Biol.* **2002**, *322*, 463.
(105) Guerois, R.; Serrano, L. *J. Mol. Biol.* **2000**, *304*, 967.
(106) Munoz, V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11311.
(107) Makarov, D. E.; Plaxco, K. W. *Protein Sci.* **2003**, *12*, 17.
(108) Garcia-Mira, M. M.; Boehringer, D.; Schmid, F. X. *J. Mol. Biol.* **2004**, *339*, 555.
(109) Grantcharova, V. P.; Riddle, D. S.; Santiago, J. V.; Baker, D. *Nat. Struct. Biol.* **1998**, *5*, 714.
(110) Gruebele, M.; Wolynes, P. G. *Nat. Struct. Biol.* **1998**, *5*, 662.
(111) Guo, W.; Lampoudi, S.; Shea, J. E. *Proteins* **2004**, *55*, 395.
(112) Klimov, D. K.; Thirumalai, D. *Proteins* **2001**, *43*, 465.
(113) Lindberg, M.; Tangrot, J.; Oliveberg, M. *Nat. Struct. Biol.* **2002**, *9*, 818.
(114) Riddle, D. S.; Grantcharova, V. P.; Santiago, J. V.; Alm, E.; Ruczinski, I. I.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 1016.
(115) Weikl, T. R.; Dill, K. A. *J. Mol. Biol.* **2003**, *332*, 953.
(116) Yi, Q.; Rajagopal, P.; Klevit, R. E.; Baker, D. *Protein Sci.* **2003**, *12*, 776.
(117) Gruebele, M. *Curr. Opin. Struct. Biol.* **2002**, *12*, 161.
(118) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.
(119) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins* **1995**, *21*, 167.
(120) Dill, K. A.; Chan, H. S. *Nat. Struct. Biol.* **1997**, *4*, 19.
(121) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. *Fold. Des.* **1996**, *1*, 441.
(122) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *Proteins* **1998**, *32*, 136.
(123) Shakhnovich, E. I. *Fold. Des.* **1998**, *3*, R108.
(124) Thirumalai, D.; Klimov, D. K. *Fold. Des.* **1998**, *3*, R112.
(125) Yang, W. Y.; Pitera, J. W.; Swope, W. C.; Gruebele, M. *J. Mol. Biol.* **2004**, *336*, 241.
(126) Burton, R. E.; Myers, J. K.; Oas, T. G. *Biochemistry* **1998**, *37*, 5337.
(127) Baldwin, R. L. *J. Biomol. NMR* **1995**, *5*, 103.
(128) Wright, C. F.; Lindorff-Larsen, K.; Randles, L. G.; Clarke, J. *Nat. Struct. Biol.* **2003**, *10*, 658.
(129) Zaidi, F. N.; Nath, U.; Udgaonkar, J. B. *Nat. Struct. Biol.* **1997**, *4*, 1016.
(130) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3948.
(131) Leeson, D. T.; Gai, F.; Rodriguez, H. M.; Gregoret, L. M.; Dyer, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2527.
(132) Martinez, J. C.; Serrano, L. *Nat. Struct. Biol.* **1999**, *6*, 1010.
(133) Grantcharova, V. P.; Riddle, D. S.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7084.
(134) Viguera, A. R.; Serrano, L. *Nat. Struct. Biol.* **1997**, *4*, 939.
(135) Grantcharova, V.; Alm, E. J.; Baker, D.; Horwich, A. L. *Curr. Opin. Struct. Biol.* **2001**, *11*, 70.
(136) Viguera, A. R.; Serrano, L.; Wilmanns, M. *Nat. Struct. Biol.* **1996**, *3*, 874.
(137) Hennecke, J.; Sebbel, P.; Glockshuber, R. *J. Mol. Biol.* **1999**, *286*, 1197.
(138) Grantcharova, V. P.; Baker, D. *J. Mol. Biol.* **2001**, *306*, 555.
(139) Otzen, D. E.; Fersht, A. R. *Biochemistry* **1998**, *37*, 8139.
(140) Maxwell, K. L.; Wildes, D.; Zarrine-Afsar, A.; De Los Rios, M. A.; Brown, A. G.; Friel, C. T.; Hedberg, L.; Horng, J. C.; Bona, D.; Miller, E. J.; Vallee-Belisle, A.; Main, E. R.; Bemporad, F.; Qiu, L.; Teilum, K.; Vu, N. D.; Edwards, A. M.; Ruczinski, I.; Poulsen, F. M.; Kragelund, B. B.; Michnick, S. W.; Chiti, F.; Bai, Y.; Hagen, S. J.; Serrano, L.; Oliveberg, M.; Raleigh, D. P.; Wittung-Stafshede, P.; Radford, S. E.; Jackson, S. E.; Sosnick, T. R.; Marqusee, S.; Davidson, A. R.; Plaxco, K. W. *Protein Sci.* **2005**, *14*, 602.

CR040431Q