doi:10.1016/j.jmb.2012.04.013 J. Mol. Biol. (2012) 420, 220–234

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

The Folding Transition State of Protein L Is Extensive with Nonnative Interactions (and Not Small and Polarized)

Tae Yeon Yoo[1], Aashish Adhikari[2], Zhen Xia[4], Tien Huynh[4], Karl F. Freed[2,3]*, Ruhong Zhou[4]* and Tobin R. Sosnick[3,5]*

[1]*Department of Physics, University of Chicago, Chicago, IL 60637, USA*
[2]*Department of Chemistry and The James Franck Institute, University of Chicago, Chicago, IL 60637, USA*
[3]*Computation Institute, University of Chicago, Chicago, IL 60637, USA*
[4]*Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA*
[5]*Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637, USA*

Received 6 February 2012; received in revised form 4 April 2012; accepted 11 April 2012
Available online 18 April 2012

Edited by C. R. Matthews

Keywords:
ϕ analysis;
ψ analysis;
metal binding;
bi-histidine;
protein folding

Progress in understanding protein folding relies heavily upon an interplay between experiment and theory. In particular, readily interpretable experimental data that can be meaningfully compared to simulations are required. According to standard mutational ϕ analysis, the transition state for Protein L contains only a single hairpin. However, we demonstrate here using ψ analysis with engineered metal ion binding sites that the transition state is extensive, containing the entire four-stranded β sheet. Underreporting of the structural content of the transition state by ϕ analysis also occurs for acyl phosphatase [Pandit, A. D., Jha, A., Freed, K. F. & Sosnick, T. R., (2006). Small proteins fold through transition states with native-like topologies. *J. Mol. Biol.* **361**, 755–770], ubiquitin [Sosnick, T. R., Dothager, R. S. & Krantz, B. A., (2004). Differences in the folding transition state of ubiquitin indicated by ϕ and ψ analyses. *Proc. Natl Acad. Sci. USA* **101**, 17377–17382] and BdpA [Baxa, M., Freed, K. F. & Sosnick, T. R., (2008). Quantifying the structural requirements of the folding transition state of protein A and other systems. *J. Mol. Biol.* **381**, 1362–1381]. The carboxy-terminal hairpin in the transition state of Protein L is found to be nonnative, a significant result that agrees with our Protein Data Bank-based backbone sampling and all-atom simulations. The nonnative character partially explains the failure of accepted experimental and native-centric computational approaches to adequately describe the transition state. Hence, caution is required even when an apparent agreement exists between experiment and theory, thus highlighting the importance of having alternative methods for characterizing transition states.

*Corresponding authors. E-mail addresses: freed@uchicago.edu; ruhongz@us.ibm.com; trsosnic@uchicago.edu.

Abbreviations used: biHis, bi-histidine; GdmCl, guanidinium chloride; MD, molecular dynamics; RCO, relative contact order; TSE, transition state ensemble; PDB, Protein Data Bank.

Introduction

The IgG binding domain of Protein L (Protein L) contains two hairpins and a central helix and has been a test bed for many experimental and theoretical studies of folding.[1–11] Mutational ϕ

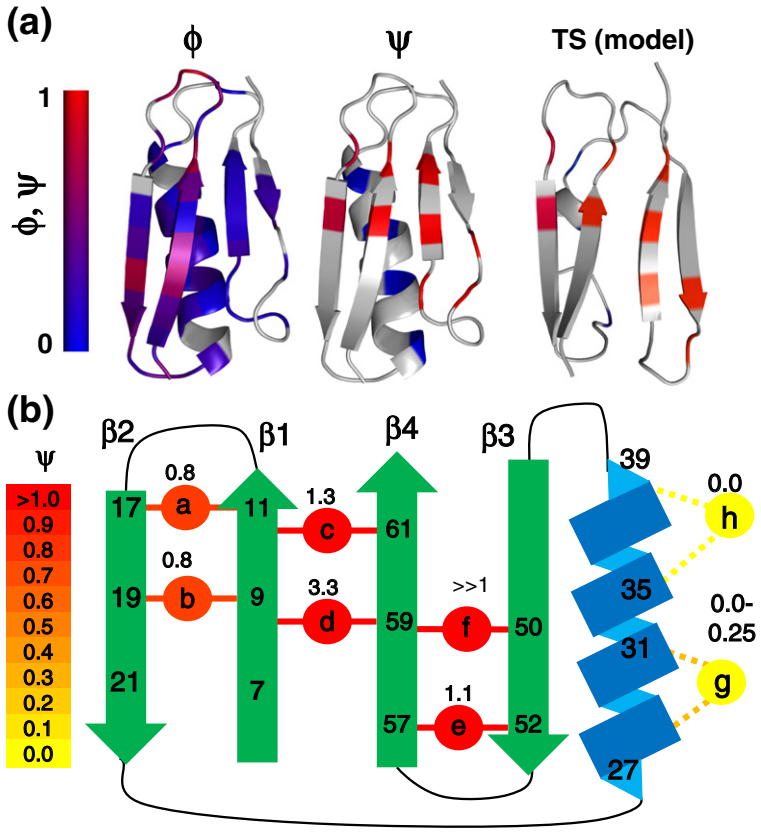

Fig. 1. BiHis sites, ψ and ϕ values and TSE for Protein L. (a) ϕ and ψ values mapped onto Protein L. The ψ are obtained from the present study, while the ϕ are from Ref. 4. When multiple mutations are studied at a particular site, more emphasis is placed on the X-to-Ala and Ala-to-Gly comparisons for core and helical mutations, respectively. The model of the TSE is created by docking the two hairpins using the *ItFix* protocols. (b) The locations of the biHis sites and ψ values are shown (each site is studied individually). Renderings created in PyMOL.

analysis experiments indicate that the folding transition state ensemble (TSE) contains only the amino-terminal hairpin[1–4] (Fig. 1). The TSE of a protein with the same α/β fold, Protein G, is also assigned by ϕ analysis to have a single hairpin, but this hairpin is located at the carboxy terminus,[12] a behavior attributed to different properties of the turn sequences.[13,14] The difference between the TSEs of these two proteins is cited as an example where the specific sequence, rather than just the protein's topology, influences the folding pathway. A variety of computational studies support this view.[5–9,11]

Despite this broad consensus, we decided to reexamine the folding behavior of Protein L because a TSE with only a single hairpin seems inordinately small. A hairpin scarcely defines Protein L's topology, yet this protein obeys the well-known correlation between folding rate and topology [relative contact order (RCO)].[15,16] Our studies of three other proteins with disparate RCOs indicate that their TSEs acquire a similar level of native topology, $RCO^{TSE} \approx 0.7 \cdot RCO^{Native}$.[17–21] If this relationship is generally applicable, it would provide a simple rationalization for the k_f–RCO correlation, as well as a constraint for possible TSE structures of other proteins.

In the case of Protein L, the presence of only a single hairpin in the TSE equates to an RCO fraction of only 25%, and even the inclusion of the helix would increase the RCO only to 40%. The TSE must minimally include long-range contacts between the amino- and carboxy-terminal strands in order to achieve an RCO fraction close to 70%. Furthermore, whereas a 1:1 relationship between hydrogen bond content and surface burial is found in the TSEs of other proteins,[22,23] the hydrogen bond content of a single hairpin is grossly inadequate to match the surface burial of the highly collapsed TSE of Protein L as determined by the denaturant dependence of the folding rates.

Here, we employ ψ analysis[24] to characterize the TSE structure of Protein L. ψ is well suited for determining the structure because the methodology directly identifies pairwise residue–residue contacts. The methodology employs bi-histidine (biHis) metal ion binding sites on the surface of protein, which are stabilized by the addition of metal ions. The ion-induced stabilization of the TSE relative to the native state is represented by the ψ value that is high if the biHis site is present in the TSE. Data for a multitude of biHis sites (individually introduced) can be used to generate structural models of the TSE analogous to the use of nuclear Overhauser enhancement distance constraints in NMR-based structure determination.

These experiments demonstrate that Protein L's TSE contains the entire four-stranded β sheet. Although the amino-terminal hairpin is native-like,

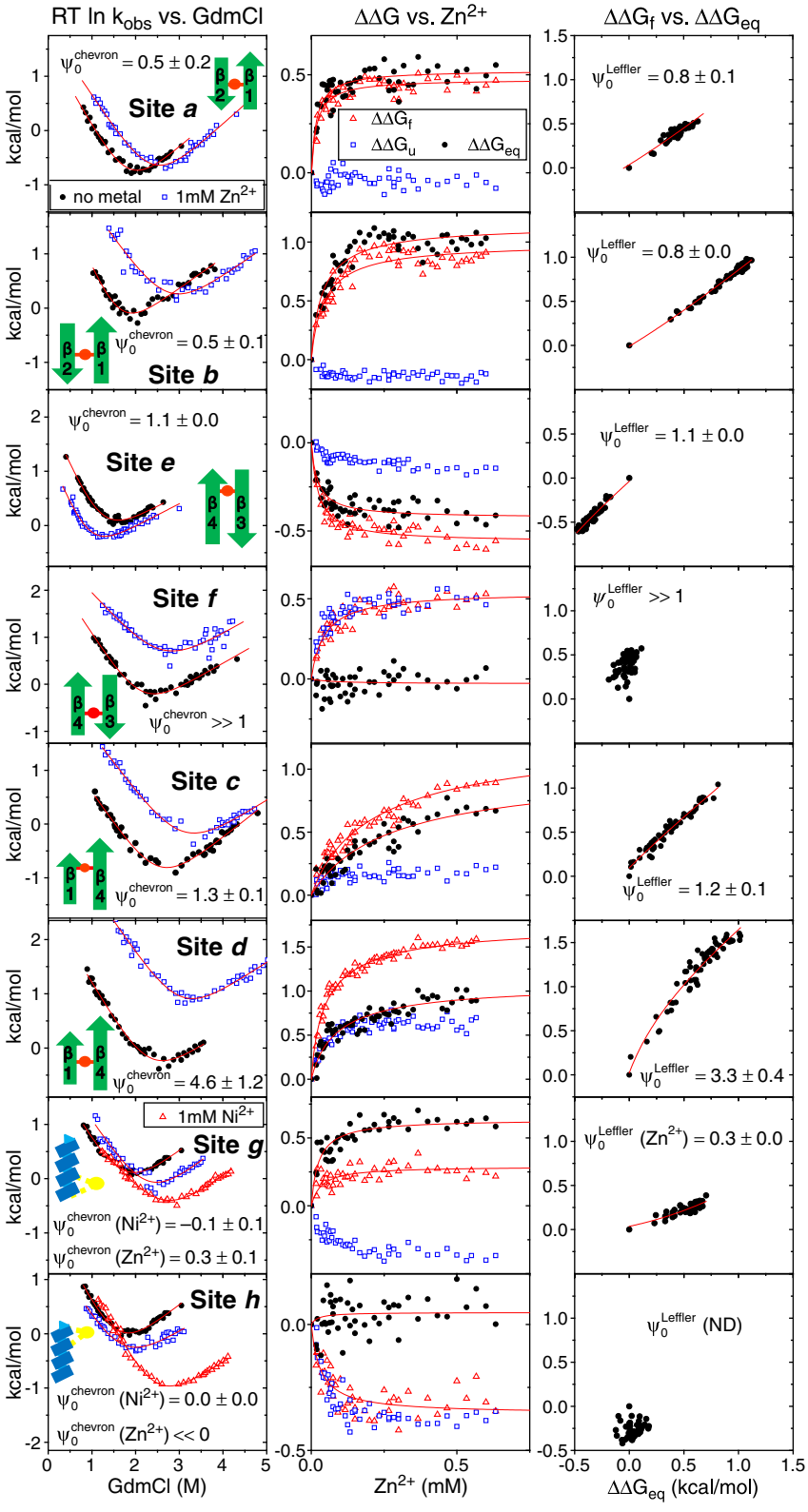

Fig. 2. Metal-dependent kinetic data. Left: The denaturant chevrons in the absence of metal ions (black filled circle), in the presence of 1 mM Zn^{2+} (blue open square) and with Ni^{2+} (red open triangle). Center: Kinetics as a function of Zn^{2+} at fixed [GdmCl]. The changes in stability $\Delta\Delta G_{eq}$ (black filled circle), folding activation energy $\Delta\Delta G_f$ (red open triangle) and unfolding activation energy $\Delta\Delta G_u$ (blue open square) plotted against Zn^{2+} concentration. Values are calculated by measuring the folding and unfolding rates at fixed low (~0.6 M) and high (~3.5–5 M) GdmCl concentrations, respectively (Table 2). The binding affinities are obtained from fitting to the Eqs. (1). Right: Corresponding Leffler plots.

the carboxy hairpin and the long-range interactions between the two hairpins have nonnative properties. We conduct simulations of the individual hairpins using our *ItFix* folding algorithm where the side chains are represented by single C^{β} atoms,[25,26] as well as all-atom explicit solvent

molecular dynamics (MD) simulations. Without invoking any knowledge on the native state, both methods indicate that the carboxy-terminal hairpin forms rapidly, but with a nonnative turn. We discuss the implications of our findings with regard to TSE topology, the accuracy of ϕ analysis and its ability to validate theoretical studies.

Results

ψ analysis[17,19,20,24,27,28] proceeds by introducing biHis metal ion binding sites at positions across the protein's surface. A total of eight sites are individually introduced into Protein L to probe the formation of the three native strand–strand pairings and the helix (Fig. 1). Upon addition of metal ions, the biHis sites stabilize strand–strand pairings or the helix because an increase in metal ion concentration stabilizes the interaction between the two histidine partners. The changes in protein's equilibrium stability and folding activation free energy ($\Delta\Delta G_{eq}$ and $\Delta\Delta G_f$, respectively) arise from the difference in metal ion binding disassociation constants (K^U, K^N and K^{TSE}) of the biHis site in the unfolded (U) state, native (N) state and TSE, as given by

$$\Delta\Delta G_{eq}\left(\left[Me^{2+}\right]\right) = RT\ln\left(1 + \left[Me^{2+}\right] / K^N\right) - RT\ln\left(1 + \left[Me^{2+}\right] / K^U\right) \quad (1a)$$

$$\Delta\Delta G_f\left(\left[Me^{2+}\right]\right) = RT\ln\left(1 + \left[Me^{2+}\right] / K^{TSE}\right) - RT\ln\left(1 + \left[Me^{2+}\right] / K^U\right) \quad (1b)$$

These ion-induced changes in free energies are used to define the ψ value, a parameter analogous to the standard mutational ϕ value, although ψ is the instantaneous change as a function of metal ion:

$$\phi = \frac{\Delta\Delta G_f}{\Delta\Delta G_{eq}}\Big|_{mutation} \quad (2a)$$

$$\psi\left(\left[Me^{2+}\right]\right) = \frac{\partial\Delta\Delta G_f}{\partial\Delta\Delta G_{eq}}\Big|_{\left[Me^{2+}\right]} \quad (2b)$$

To remove any potential artifacts related to the alteration of the folding behavior by the metal ion binding, we obtain a "ψ_0" value by evaluating ψ in the limit of $[Me^{2+}] \to 0$. The ψ_0 thus reflects the intrinsic degree of contact formation in the TSE when metal ions are absent. Hence, ψ analysis provides information on the conformation of the TSE prior to any ion-induced perturbation. In particular, ψ analysis probes the metal ion binding affinity of the two histidines in the TSE; for example, if the two residues are pre-positioned to bind ions tighter in the TSE than in the native state, this stabilization arises because the TSE adopts a nonnative conformation, not because the ions induce a conformational change. This ability to identify the intrinsic folding behavior in the absence of the metal ion perturbation lacks a direct counterpart with ϕ analysis, which typically does not address the possible consequences of even sizable substitutions on altering the properties of the TSE.

ψ_0 values of zero or unity indicate that the biHis site has the ion binding affinity found in the unfolded or native site, respectively. These values are interpreted as implying that the biHis site is absent or native-like in the TS, respectively. A fractional ψ value indicates that either the biHis site is native-like in a subpopulation of the TSE or the site contains nonnative binding affinity in the entire TSE (e.g., a distorted site with less favorable binding geometry or a flexible site that must be restricted prior to ion binding), or some combination thereof.[18,28] A more thorough description of ψ analysis is provided in the appendix.

The folding properties and ψ_0 value of each biHis variant are measured using two independent approaches. First, the dependence of the folding and unfolding rates on the concentration of guanidinium chloride (GdmCl) {"chevron analysis" with plots $RT\ln(k_{observed})$ *versus* [denaturant]} is obtained in the absence and presence of 1 mM zinc or nickel ions at 22 °C, pH 7.5 (Fig. 2, left panels). Second, kinetic data are taken at dozens of different metal concentrations (Fig. 2, center and right panels) under strongly folding (~0.6 M GdmCl) and strongly unfolding (~3.5–5 M GdmCl, depending upon variant) conditions.

Using each of the two classes of data, we evaluate the ψ_0 value for each site using an equation derived from Eqs. (1) and (2):

$$\psi_0 = \left[e^{\Delta\Delta G_f} - 1\right] / \left[e^{\Delta\Delta G_{eq}} - 1\right] \quad (3)$$

A $\psi_0^{chevron}$ value is evaluated from the changes in the folding and unfolding rates (i.e., the shift in the arms of the chevron plots) arising from the addition of 1 mM metal ion. The magnitude of the shifts generates a single ($\Delta\Delta G_{eq}$,$\Delta\Delta G_f$) pair. This pair is sufficient to determine ψ_0 using Eq. (3). This procedure is analogous to the method for calculating ϕ by comparing the chevrons for the wild-type and mutant proteins.

The second, independent determination of ψ_0 (termed $\psi_0^{Leffler}$), is obtained from the fit of the $\Delta\Delta G_{eq}$ *versus* $\Delta\Delta G_f$ data, as presented in the Leffler plot (Fig. 2, right panel). Here, the multitude of ($\Delta\Delta G_{eq}$,$\Delta\Delta G_f$) points obtained using the kinetic data taken at dozens of metal ion concentrations are fit using a rearranged form of Eq. (3): $\Delta\Delta G_f = RT\ln\left((1 - \psi_0) + \psi_0 e^{\Delta\Delta G_{eq}/RT}\right)$. In addition,

the metal ion dependence of $\Delta\Delta G_{eq}$ and $\Delta\Delta G_f$ (Fig. 2, central panels) and Eqs. (1) are used to individually determine each of the dissociation constants K^U, K^N and K^{TSE} (Table 1).

Structure in Protein L's TSE

Six biHis sites are located on the sheets, and two sites are situated along the helix. Sites *a* & *b* and *e* & *f* are located on the amino and carboxy hairpins, respectively. Sites *c* & *d* lie between the amino and carboxy strands and connect the two hairpins. Sites *g* and *h* are located in *i*, *i*+4 positions along the sole helix.

The sensitivity of folding rates to metal ion concentration indicated that Protein L's TSE had all six biHis sites on the β sheet at least partially formed, while the helix was largely absent (Fig. 2 and Tables 1 and 2). For sites *a* and *b* on the amino hairpin, the histidines pairs had a near-native binding affinity with $\psi_0^{Leffler} = 0.77 \pm 0.07$ and 0.75 ± 0.03, while $\psi_0^{chevron}$ was slightly lower (0.51 ± 0.17 and 0.48 ± 0.07, respectively). These data indicated that the hairpin was formed in the TSE, although with the biHis site having slightly weaker ion binding affinity in the TSE than in the native state. This interpretation is consistent with the high values of ϕ measured for mutations throughout the hairpin.[4]

However, the folding behavior for sites *e* and *f* and *c* and *d* on the carboxy hairpin and between the two hairpins, respectively, were indicative of a nonnative arrangement of the strands. The equilibrium stability for the site *e* and *f* variants decreased and remained unchanged, respectively. Relaxation rates were Zn^{2+} dependent for both positions. This unusual behavior indicated that the metal ion binding affinity in the native state was weaker than or comparable to that in the unfolded state. The relative binding affinities were $K^{TSE}/K^N = 0.48 \pm 0.07$ and 0.95 ± 0.23 for sites *e* and *f*, respectively. The stronger metal ion binding affinity in the unfolded state for site *e*, which is positioned closer to the turn, can be explained by the presence of structure in the denatured state having a biHis arrangement that binds zinc ions stronger than the biHis site in the native state.

The folding rate of site *e* and, to a much smaller degree, the unfolding rate were decelerated upon addition of zinc ions. The resulting ψ was near unity, $\psi_0^{Leffler} = 1.13 \pm 0.03$, indicating that the carboxy-terminal hairpin is formed in the TSE.

Site *f*, located at the distal end of the carboxy hairpin, presented folding and unfolding rates that were equally accelerated upon addition of zinc ions (Fig. 2). The metal ion binding affinity in the TSE was 2.6-fold stronger than in the native or unfolded state. The denaturant chevron was shifted upward with the vertex at the same GdmCl concentration. Hence, only the relative barrier height decreased upon addition of metal ions, while the relative depth of the unfolded and native-state wells on either side remained unchanged.

ψ_0 itself was ill-determined because the change in equilibrium stability was near zero. Nevertheless, the stronger metal ion binding affinity in the TSE than in either the U or N states indicated that, in the TSE, the two histidines formed a binding site with a preferred orientation or distance. Hence, the carboxy-terminal hairpin was also present in the TSE.

Upon addition of metal ions for sites *c* and *d* that connect the two hairpins, the activation energy for folding was affected more than the equilibrium stability (Fig. 2). Consequently, ψ_0 exceeded unity, with $\psi_0^{Leffler} = 1.24 \pm 0.07$ and 3.33 ± 0.40 for sites *c* and *d*, respectively. The metal ion binding affinities in the TSE were 1.5-fold and 3-fold tighter than in the native state, respectively. Hence, the two hairpins were in close proximity in the TSE, but the two histidines resided in a configuration where they could bind metal ions tighter than the native orientation. We reiterate that because ψ_0 was calculated at the limit of zero metal ion concentration (e.g., as the single free parameter in a fit to the data in the Leffler plot), the nonnative biHis geometry was not induced by the metal binding.

The application of ψ analysis to the helical sites *g* and *h* indicated that the helix was largely absent in the TSE, consistent with deductions from ϕ

Table 1. ψ values and Zn^{2+} dissociation constants

Site	Mutation	$\psi_0^{Leffler}$	$\psi_0^{chevron}$	K^U (μM)	K^N (μM)	K^{TSE} (μM)	K^U/K^N	K^{TSE}/K^N
a	I11H T17H (β1-β2)	0.77 ± 0.07	0.51 ± 0.17	26.5 ± 3.3	10.8 ± 1.2	11.7 ± 1.3	2.44 ± 0.42	1.08 ± 0.18
b	N9H T19H (β1-β2)	0.75 ± 0.03	0.48 ± 0.07	105 ± 11	14.9 ± 1.1	19.3 ± 1.5	7.02 ± 0.90	1.29 ± 0.14
c	I11H K61H (β1-β4)	1.24 ± 0.07	1.28 ± 0.10	697 ± 107	122 ± 9	79.3 ± 4.8	5.69 ± 0.96	0.65 ± 0.06
d	N9H N59H (β1-β4)	3.33 ± 0.40	4.64 ± 1.24	232 ± 21	36.8 ± 2.0	11.8 ± 0.6	6.30 ± 0.66	0.32 ± 0.02
e	A52H T57H (β3-β4)	1.13 ± 0.03	1.13 ± 0.02	15.2 ± 1.4	31.4 ± 3.3	39.7 ± 4.3	0.48 ± 0.07	1.26 ± 0.19
f	D50H N59H (β3-β4)	≫ 1[a]	≫ 1[a]	71.5 ± 12.0	75.3 ± 12.9	28.6 ± 4.0	0.95 ± 0.23	0.38 ± 0.08
g	K28H E32H (helix)	0.26 ± 0.03	0.25 ± 0.09 (−0.05 ± 0.05)[b]	43.1 ± 6.5	14.5 ± 1.9	26.3 ± 3.7	2.97 ± 0.59	1.81 ± 0.35
h	A35H T39H (helix)	≪ 0[a]	(0.02 ± 0.004)[b]	23.3 ± 5.6	21.4 ± 5.1	42.6 ± 11.4	1.09 ± 0.37	1.99 ± 0.71

[a] The metal-induced stabilization is too small to provide a well-defined ψ value.
[b] Values in parentheses are obtained using Ni^{2+} ions.

Table 2. Equilibrium and kinetic parameters for divalent metal ion binding

Site	Mutation (location)	$\Delta\Delta G_{mut}$	$\Delta\Delta G_{eq}^{equil}$ ($\Delta\Delta G_{eq}^{kin}$)[a]	$\Delta\Delta G_f^{\ddagger}$ ($\Delta G_f^{\ddagger}$)	m_0 (m_0^{Me})	m_f/m_0 (m_f^{Me}/m_0^{ME})	$\psi_0^{Leffler}$ ($\psi_0^{chevron}$)[b]	Metal
WT	NA	NA	NA	NA	1.97 ± 0.08	0.73 ± 0.05	NA	NA
				(1.51 ± 0.06)	(NA)	(NA)		
a	I11H	0.14 ± 0.02	0.88 ± 0.03	0.46 ± 0.17	2.11 ± 0.11	0.67 ± 0.05	0.77 ± 0.07	Zn
	T17H		(0.70 ± 0.18)[c]	(0.79 ± 0.05)	(1.92 ± 0.06)	(0.63 ± 0.03)	(0.51 ± 0.17)	
	(β1-β2)							
b	N9H	−0.32 ± 0.02	1.31 ± 0.13	0.77 ± 0.22	2.38 ± 0.22	0.76 ± 0.11	0.75 ± 0.03	Zn
	T19H		(1.11 ± 0.24)	(1.64 ± 0.16)	(1.68 ± 0.12)	(0.67 ± 0.08)	(0.48 ± 0.07)	
	(β1-β2)							
c	I11H K61H	−0.45 ± 0.03	ND	1.03 ± 0.10	1.81 ± 0.07	0.65 ± 0.04	1.24 ± 0.07	Zn
	(β1-β4)		(0.92 ± 0.10)	(1.18 ± 0.06)	(1.58 ± 0.06)	(0.68 ± 0.04)	(1.28 ± 0.10)	
d	N9H N59H	0.53 ± 0.02	1.03 ± 0.15	1.63 ± 0.25	1.87 ± 0.11	0.73 ± 0.06	3.33 ± 0.40	Zn
	(β1-β4)		(0.85 ± 0.16)	(1.88 ± 0.05)	(1.70 ± 0.07)	(0.69 ± 0.04)	(4.64 ± 1.24)	
e	A52H T57H	−0.88 ± 0.02	−0.60 ± 0.07	−0.76 ± 0.05	2.22 ± 0.07	0.81 ± 0.04	1.13 ± 0.03	Zn
	(β3-β4)		(−0.60 ± 0.03)[d]	(1.11 ± 0.02)	(2.38 ± 0.11)	(0.82 ± 0.06)	(1.13 ± 0.02)[e]	
f	D50H N59H	0.64 ± 0.02	0.21 ± 0.17	0.66 ± 2.47	1.96 ± 0.13	0.73 ± 0.08	ND[f]	Zn
	(β3-β4)		(0.03 ± 0.12)	(1.79 ± 0.10)	(1.54 ± 0.11)	(0.65 ± 0.08)		
g	K28H E32H	−0.46 ± 0.02	1.10 ± 0.04	0.29 ± 0.16	1.92 ± 0.14	0.69 ± 0.08	0.26 ± 0.03	Zn
	(helix)		(0.76 ± 0.24)	(1.37 ± 0.05)	(1.81 ± 0.11)	(0.66 ± 0.06)	(0.25 ± 0.09)	
			1.20 ± 0.12	−0.13 ± 0.16	1.92 ± 0.14	0.69 ± 0.08	NA[g]	Ni
			(0.92 ± 0.22)	(1.37 ± 0.05)	(1.57 ± 0.09)	(0.61 ± 0.06)	(−0.05 ± 0.05)	
h	A35H T39H	0.30 ± 0.01	0.48 ± 0.11	−0.38 ± 0.65	2.14 ± 0.12	0.71 ± 0.06	ND[f]	Zn
	(helix)		(0.33 ± 0.20)	(1.32 ± 0.05)	(1.78 ± 0.13)	(0.73 ± 0.08)		
			1.75 ± 0.07	0.28 ± 0.08	2.14 ± 0.12	0.71 ± 0.06	NA[g]	Ni
			(2.01 ± 0.16)	(1.32 ± 0.05)	(1.98 ± 0.06)	(0.75 ± 0.04)	(0.02 ± 0.004)	

Units are expressed as kilocalories per mole (kcal·mol^{-1}) (free energies) or kilocalories per mole per molar (kcal·mol^{-1}·M^{-1}) (*m*-values). NA, not applicable. ND, not determined.

[a] $\Delta\Delta G_{eq}^{equil}$ is the metal-induced stabilization determined by 280 nm fluorescence *versus* GdmCl concentration curve from equilibrium denaturation measurement unless mentioned otherwise. $\Delta\Delta G_{eq}^{kin}$ is the metal-induced stabilization obtained from the simultaneous fit of double chevrons. To minimize extrapolation errors, we calculated changes in stabilities at 0.5 and 4.5 M GdmCl, and we obtain them by simultaneously fitting two chevrons in the absence and presence of 1 mM metal ion, with the parameter of interest being one of the fitting parameters.

[b] $\psi_0^{Leffler}$ and $\psi_0^{chevron}$ are the ψ_0 values obtained from double-chevron analysis and the fit of a Leffler plot, respectively.

[c] $\Delta\Delta G_{eq}^{equil}$ is determined by $\Delta\Delta G_{eq}^{equil} = \Delta C_m \times m_0^{No\ metal}$ where ΔC_m is the metal-induced change in midpoint, which is the GdmCl concentration where the folded and unfolded populations are the same. $m_0^{No\ metal}$ is the m_0 obtained in the absence of metal ion.

[d] $\Delta\Delta G_{eq}^{equil}$ is determined by equilibrium denaturation using circular dichroism measurements at $\lambda = 222$ nm instead of fluorescence measurement. The m_0 value was shared when equilibrium denaturation data in the absence and presence of metal ion were fitted simultaneously.

[e] The slope of unfolding arm m_u was shared when two chevrons were simultaneously fitted.

[f] Chevron shifts vertically upon addition of metal ion. The metal-induced stabilization is too small to provide a well-defined ψ value.

[g] Unfolding arm has two phases of which the fast phase and the slow phase exchange their relative amplitudes as the nickel ion concentration increases while the rates themselves are constant over nickel ion concentration. The ψ_0 value could not be obtained from the Leffler plot because $\Delta\Delta G_{eq}$ cannot be determined at the intermediate nickel ion concentration where the relative amplitude of the two phases is poorly determined.

analysis.[3,4] Ni^{2+} titration only affected the unfolding chevron arm, and $\psi_0^{chevron}$ vanished for both sites (Fig. 2). Titrating the site *g* variant with Zn^{2+} ions yielded $\psi_0^{Leffler} = 0.26 \pm 0.03$. The origin of the minor difference between the ψ values obtained with the two metals ions for site *g* was unclear. The greater change in stability with Ni^{2+} (the chevron plot shifted farther to the right) suggested that the Ni^{2+} data were more reliable.

For the site *h* variant in Zn^{2+}, both folding and unfolding rates mildly *decreased*, and the change in stability was near zero: $\Delta\Delta G_{eq} = 0.33 \pm 0.20$ kcal mol^{-1}. This unexpected downward shift in the chevron plot rendered the $\psi_0^{chevron}$ ill-determined. Although the Ni^{2+} data implied the lack of the helix in the TSE, the Zn^{2+} data for the site *h* variant indicated that the TSE bound metal ions more weakly than either the unfolded or folded states, for example, $K^{TSE}/K^{N} = 1.99 \pm 0.71$. Potentially, this part of the chain adopted a non-helical arrangement in the TSE with an extended backbone geometry in which the histidines were located farther apart and with weaker ion binding affinity than in either the unfolded state (where transient helix formation could occur) or the native state (which is helical).

The kinetics of the two helical sites *g* and *h* yielded biphasic unfolding traces in Ni^{2+}, with one phase being ~10-fold slower (Fig. 3). Both rates depended on the GdmCl concentration, with the same slope as the unfolding arm of the chevron plots observed for Protein L. The amplitudes of the fast and slow phases, however, depended on metal ion concentration. The faster unfolding rate matched the unfolding rate in the absence of metal ions, while the amplitude of the slower phase increased with the nickel ion concentration.

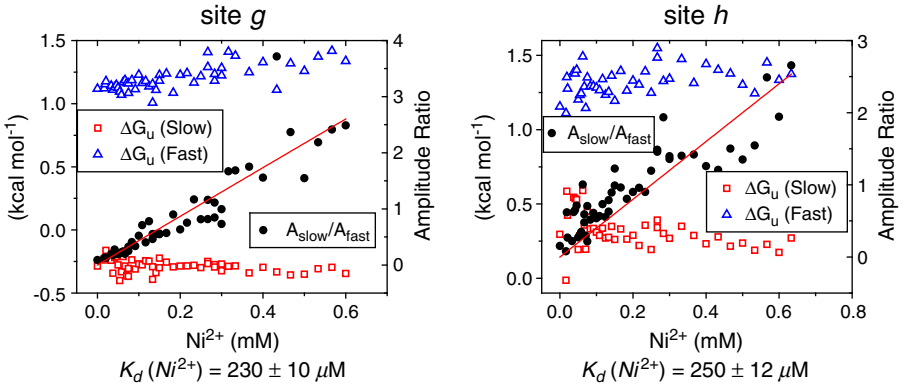

Fig. 3. Unfolding rates and amplitude dependence on metal concentration for helical sites *g* and *h*. Red line is a fit to the ratio of the two amplitudes $[A_{fast}]/[A_{slow}]=[Me^{2+}]/K_d$.

This behavior suggested that the metal ion binding equilibrium at low metal concentration was established slower than the unfolding time constant. The amplitudes of the slower and faster unfolding phases represented the fractions of the metal ion-bound and ion-unbound populations, respectively, present at the initiation of the unfolding reaction. Their relative populations reflected the ion binding affinity in the native state: [bound]/[unbound]$=[Me^{2+}]/K^N$. Regardless, the 10-fold slower unfolding rate for the metal ion-bound form was consistent with the loss of helical structure in the route from the native to the TSE.

Hairpin simulations and TSE modeling

To gain further insight into the nonnative ψ values, we conducted simulations of the individual hairpins using both our homology-free *ItFix* folding algorithm[25,26] and standard, explicit solvent MD simulations. Both methods do not invoke any knowledge on the native state, and they concurred that the carboxy-terminal hairpin forms with a nonnative turn geometry. This result was a consequence of the native turn having three consecutive residues with positive backbone ϕ dihedral angles (Fig. 4a).

The *ItFix* MCSA (*M*onte *C*arlo *s*imulated *a*nnealing) simulations represent each side chain with a single C^β atom. The conformational search space is restricted by iteratively fixing 2° structure assignments of certain portions of the sequence after incorporating the influence of 3° context, thereby coupling secondary and tertiary structure determinations. This rapid algorithm can generate accurate predictions of both secondary and tertiary structures without relying on known structures, templates or fragments.[25,26] The method has been validated in CASP8 and CASP9 and was ranked as one of the best groups in the CASP9 refinement category that involves improving template-based models so that they can function as molecular replacement models to solve the phase problem for crystallographic structure determination.[29]

The *ItFix* move set involves changes only in a single residue's ϕ,ψ backbone dihedral angles (i.e., not a fragment insertion method). The dihedral angles are derived from a Protein Data Bank (PDB)-based coil library lacking helices and sheets, and the angles depend on the amino acid type of the residue and each flanking residue. The energy function is the sum of our orientation-dependent DOPE-PW statistical potential[26] plus a burial term based on the number of heavy atoms in an 11-Å *hemisphere* in the direction of the C^α–C^β vector for each amino acid type.[29]

Energy minimization in each successive round proceeds with an energy function that also includes the distance constraints derived from the average residue–residue contact and hydrogen bond maps of the structures with the 25% lowest energies from the previous round. Hence, information concerning both 2° and 3° structures learned in prior rounds is carried over to the next round. This iterative fixing protocol mimics the sequential stabilization process observed in the folding of real proteins wherein steps represent the building of new portions on top of existing structures.[18,30]

The *ItFix* algorithm predicted that the amino hairpin adopts a near-native conformation (Figs. 4 and 5). However, the carboxy hairpin folded into a nonnative hairpin with a two-residue registry shift in the hydrogen bonding partners. This hairpin folded in a single *ItFix* round, with 98% of the structures containing hydrogen bonds between the three nonnative residues. In contrast, the amino-terminal hairpin required two rounds of folding to obtain a native hairpin conformation. After the first and second rounds, 19% and 81% of the native hydrogen bonds were formed, respectively. Both predicted hairpins were Type 1 β turns, which have dihedral angles that correspond to the dominant

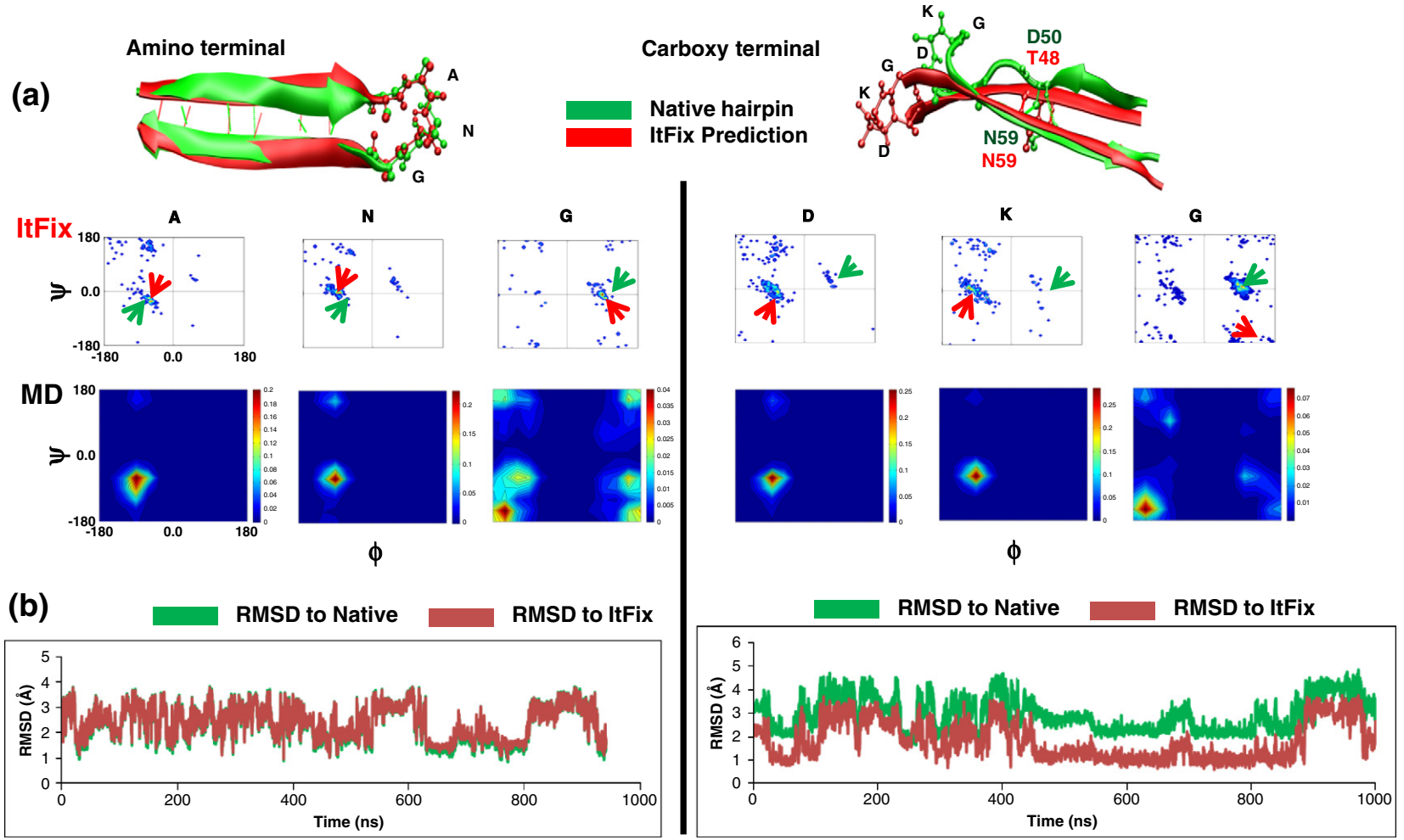

Fig. 4. Simulations of the β hairpins. (a) *ItFix* predictions overlaid on the native hairpins. Ramachandran distributions during the simulations displayed for both *ItFix* and all-atom MD simulations. The native (green arrow) and predicted (red arrow) torsional angles are depicted. (b) The RMSD values of the six turn residues in the MD simulations to the native and to the *ItFix* predicted turns.

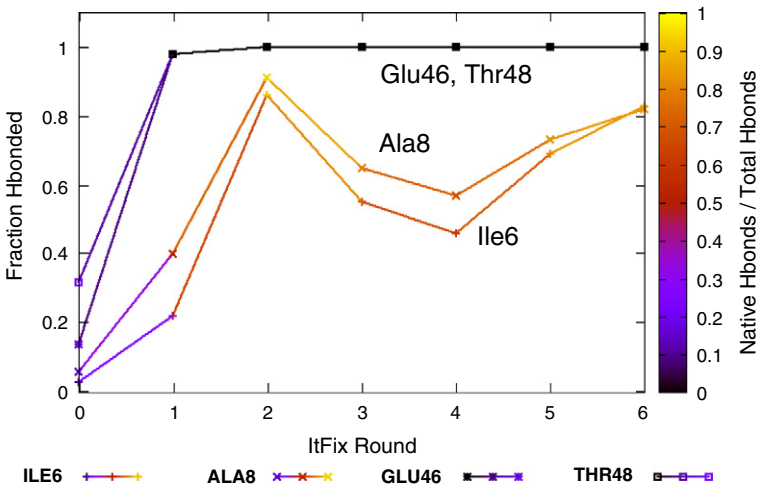

Fig. 5. Rate of β hairpin formation. The carboxy-terminal hairpin forms in fewer rounds than the amino-terminal hairpin. The carboxy hairpin folds in a single round, with 98% of the structures containing hydrogen bonds between the three nonnative amino acid partners. In contrast, the amino-terminal hairpin requires two rounds of folding to obtain a native hairpin conformation. After the first and second rounds, 19% and 81% of the native hydrogen bonds are formed, respectively. The *y*-axis presents the fraction of the maximum number of hydrogen bonds for the hairpins. The color intensity reflects the native character (0 = nonnative and 1 = all native).

dihedral angles in the PDB-based sampling library. The nonnative character of the carboxy hairpin emerged because the native ϕ angles are positive and unfavorable for turn residues D53 and K54 and, thus, were only infrequently sampled during the *ItFix* simulations.

Over the course of the microsecond all-atom MD simulations, both hairpins sampled a variety of conformations, with an RMSD from the native state for the six turn residues fluctuating between 1 and 5 Å (Fig. 4b). However, the amino hairpin had RMSD values similar to those of the native and the *ItFix*-predicted turns, consistent with their joint RMSD being only 1 Å. Further, the MD simulation found the amino turn within 1 Å of the native turn for about 20% of the trajectory. In contrast, the RMSD of the carboxy hairpin to the native turn remained above 2 Å. For half of the MD trajectory, the RMSD relative to the *ItFix* prediction was closer: ~1 Å. Moreover, the dihedral angles of the turn

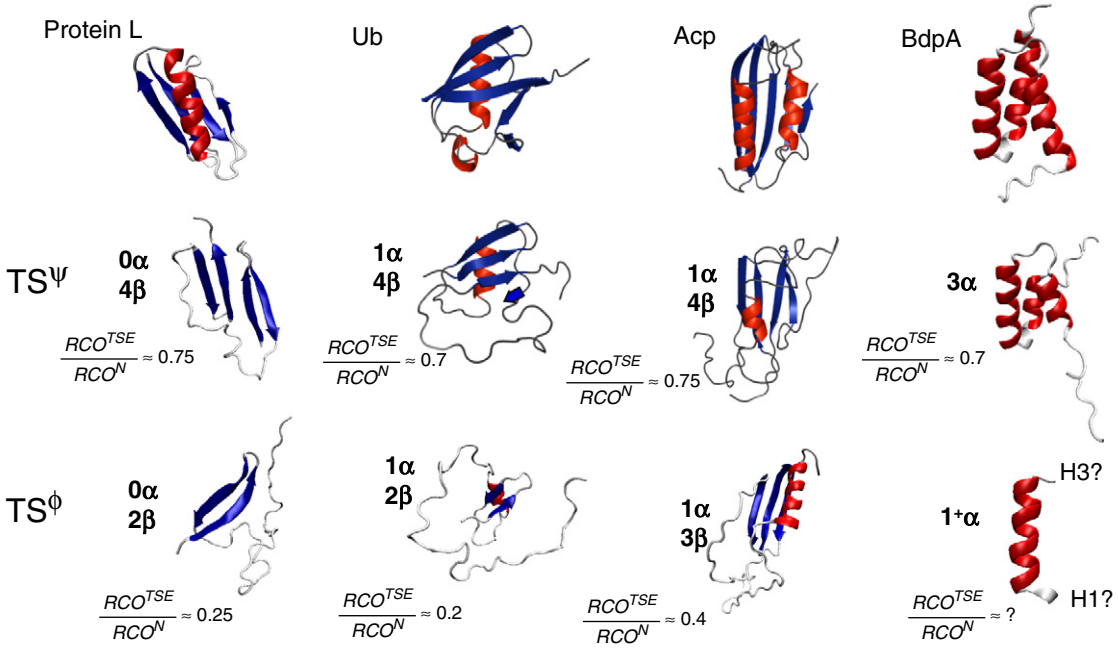

Fig. 6. TSEs determined by ψ analysis are more ordered than those identified by ϕ analysis. Although the ϕ-determined TSE depicted for Acp includes three β strands, only the presence of a hydrophobic core is indicated, and the presence of native-like strand–strand interactions is indeterminate.[51] Further, the model shown accounts for the remeasured near-zero ϕ value on F94 indicating that β5 is absent in Acp's TSE1. For BdpA, the level of the participation of helices H1 and H3 in the TSE is unclear according to ϕ analysis;[62] the model depicted reflects a combination of data from ψ and ϕ analyses and H/D kinetic isotope effects.[17]

residues from the MD simulation lie in the same basins as the *ItFix* predictions. In summary, both computational methods were consistent with the experimental findings of a native-like geometry for the amino hairpin but a nonnative geometry for the carboxy hairpin.

Models of the TSE were generated using the *ItFix* algorithm by docking the native-like amino-terminal hairpin against the nonnative carboxy hairpin (Fig. 6). Docking was achieved by introducing an additional interaction term between the terminal strands and employing single ϕ_i,ψ_i pivot moves of the unstructured residues between the two hairpins. Ten such structures were selected and further refined in a second MCSA round using our "double crank" local move set that features compensating ψ_{i-1},ϕ_i counter-rotations of the neighboring residues[29,31] (Fig. 1).

A considerable extent of Protein L's native-state topology is formed in these TSE models. After insertion of side chains using SCWRL,[32] their RCO is 73 ± 5% of the native value. The fraction of buried surface in the TSE models, relative to the native state and normalized to an unfolded state ensemble,[33] is close to the fraction of hydrogen bonds formed in the TSE, approximately 55% and 50%, respectively.

Discussion

The present study is motivated by the belief that a TSE for Protein L containing only a single hairpin, as suggested by φ analysis,[1–4] is unreasonably small because it scarcely defines the protein's topology or has enough hydrogen-bonded structure to be commensurate with the observed degree of surface burial. We have applied ψ analysis and demonstrated that the TSE is extensive, containing the entire β sheet network along with some nonnative structure associated with the carboxy hairpin. According to our simulations, this hairpin adopts a nonnative registry by virtue of the highly unfavorable dihedral angles in the native turn. The same nonnative registry is observed in both the PDB-based backbone sampling *ItFix* algorithm and the all-atom simulations. These two methods also concur on the native-like geometry of the amino hairpin, consistent with the experimental ψ data.

These findings have extensive implications concerning the role of nonnative structure, the relationship between TSE topology and folding rates and the malleability and multiplicity of TSs and whether mutational φ analysis, which has been the primary method for comparing experiment and theory, is sufficiently reliable that agreement between the two is an adequate validation of both approaches.

We observe six values of ψ_0 equal to unity or larger for biHis sites extending across the four β strands and two near-zero ψ_0 values on the helical sites. This pattern indicates that the TSE contains the entire β sheet network but minimal helical structure. The ψ_0 values greatly exceed unity for sites situated across strands β1-β4 and β3-β4 ($\psi_0^{\text{site } d} = 3.3 \pm 0.4$ and $\psi_0^{\text{site } f} \gg 1$), a behavior indicating that, in the TS, the biHis site has a geometry with tighter metal ion affinity than the site experiences in the native state. Taking further advantage of ψ analysis' ability to individually determine the metal ion binding affinities in the unfolded, transition and native states, we identified the presence of residual structure in the denatured state for the turn region of the carboxy-terminal hairpin. We emphasize that ψ_0 is the limiting ψ value in the absence of metal ions. Therefore, these properties are intrinsic to the folding behavior of Protein L and are not artifacts induced by metal ion binding.

The absence of the large helix in the TSE is attributable to its sequence having a low average intrinsic helicity: <2%.[34] In addition, all three hydrophobic residues on the buried helical face, which docks against the β sheet, are alanines rather than larger hydrophobic residues. The alanines' low hydrophobicity reduces the driving force for helix–sheet association.

Evidence that the amino acid sequence, rather than topology, can control the structure of the TSE has been found in experimental φ analysis and computational studies comparing the folding behavior of Protein L and Protein G, two proteins with the same α/β fold.[1–14] The possibility of different sequences having alternative TSEs with one hairpin or the other formed implies that a hybrid sequence could fold with significant flux going through two structurally disjoint TSEs. However, our present finding that Protein L's TSE contains both hairpins currently precludes the use of the Protein L/Protein G comparison as evidence either for sequence altering the TSE structure or for the possibility of structurally disjoint TSEs.

Our ψ-based models of the TSE (Fig. 6) are parsimonious with prior data. The fraction of the surface buried in the TSE models is close to the fraction of hydrogen bonds formed in the TSE. These findings are consistent with kinetic isotope studies that indicate the presence of a commensurate level of surface burial and hydrogen bond formation in the TSE.[22,23] The underlying principle is that hydrophobic association leads to partial backbone desolvation that can be offset by protein–protein hydrogen bonding.

A considerable extent of Protein L's topology is formed in the TSE. Our TSE models have $\text{RCO}^{\text{TSE}} \approx 0.7 \cdot \text{RCO}^{\text{Native}}$, in agreement with our prior ψ studies for ubiquitin,[18,20] acyl phosphatase[17,19] and the B domain of Protein A.[17] Because these four proteins have native RCO values that span the range observed for two state proteins, the 70% value is likely to be generalizable to other proteins that obey the well-known RCO–k_f trend.[15]

Similar relationships emerge from other studies[35] with certain Gō-based models producing a CO at the 60–80% level.[36,37] However, their TSE structures generally correspond to a uniformly "expanded version" of the native structure,[36] a finding that seems inconsistent with ψ data (many ψ values are either zero or near unity). The use of data from ϕ analysis produces an RCO^{TSE} fraction closer to 50%,[38] supporting the contention that ϕ analysis can underreport the structural content of the TSE.

The ϕ values predicted by Gō models exhibit mixed agreement with experimental ϕ values for Protein L[5] and some other proteins.[39–43] The TSE structures for Protein L from the Gō models typically contain just the amino hairpin and the helix,[5–7,9–11] but none correctly reproduces the ψ-determined TSE containing only the four strands. The closer agreement between the Gō models and ϕ analysis may partially be a consequence of their shared native-like biases.

The inability of the Gō-type simulations to correctly predict the four stands in the TSE of Protein L is likely due to both the presence of nonnative interactions and the inherent difficulty of correctly balancing the energies associated with different sets of contacts and backbone geometries. Small errors in the energy function or the lack of explicit hydrogen bonds and backbone ϕ,ψ dihedral angles can greatly impact the order in which structure forms and the location of the TSE on the reaction surface. These issues contribute to the inability of nearly all methods to accurately describe the TSE structure of a Protein L, as well as of the B domain of Protein A.[17]

Comparison to ϕ analysis

Mutational ϕ analysis is the most accepted method for characterizing TSEs, developing models for folding and validating theoretical approaches.[5,39–41,43] However, the present ψ analysis findings of an extensive TSE in Protein L significantly differ with that generated based on ϕ analysis. The ϕ analysis method indicates that Protein L's TSE contains only the amino hairpin[1–4,13] (Fig. 1). Seven sites on this hairpin yield $\phi > 0.6$ (although another five positions have ϕ below 0.31). Six positions on the carboxy-terminal hairpin produce much lower average ϕ of 0.13, while the values at two other positions are slightly higher ($\phi^{T48A} = 0.26$ and $\phi^{V49A} = 0.31$). Similarly low ϕ are found on the helical sites.

The primary differences between ψ and ϕ analyses arise because the former directly probes residue–residue contacts between two known partners, whereas the latter reflects energetic perturbations upon mutation. These perturbations may be the consequence of a combination of factors, including changes in the local side-chain environment and backbone dihedral propensities. In ψ analysis, the binding of increasing concentrations of ions to the biHis site produces a nearly continuous increase in the stability of TSE structures that contain the binding site. Hence, the stability is perturbed yet accomplished in an isosteric and isochemical manner. The resulting series of data can be justifiably combined, and the ψ_0 value can be extracted as devoid of any perturbation due to ion binding. The ability of eliminating the influence of perturbations may be inaccessible to traditional mutation studies where the perturbation can arise from multiple sources, including changes in backbone propensities as well as indeterminate nonlocal interactions.

These differences become critical for Protein L for two reasons, the nonnative character of the carboxy hairpin and the exposure of the β sheet's hydrophobic face in the TSE. The 2-aa register shift in the carboxy hairpin indicated by the simulations results in nonnative contacts along this hairpin and nonnative dihedral angles in the turn. Consequently, the energetic perturbation realized in the TSE likely is smaller than in the less accommodating native state, and ϕ therefore becomes small and mistakenly identifies this hairpin as being absent in the TSE.

The second issue arises because the otherwise buried side chains on the hydrophobic face of the sheet are solvent exposed in the TSE due to the absence of the helix. Consequently, the energetic penalty for the truncation of the side chain for the residues on the inner face of hairpins, for example, imparted by an alanine substitution, is diminished in the TSE relative to the native state, even though the residue is in a hydrogen-bonded β structure. This analysis provides an explanation for the low-to-moderate ϕ values for the native-like amino hairpin.[1–4,13] The issue is generally relevant whenever ϕ analysis is applied at any position that is more exposed in the TSE than in the native state.

Overall, these considerations support the contention that ϕ analysis can underestimate or misrepresent the structural content of the TSE[18,24,27,28,44] due to chain relaxation and accommodation or to nonnative interactions.[45,46] For example, a residue in fyn SH3 with a helical conformation in the native state adopts a β conformation in the TSE despite having a high canonical ϕ value of 0.7.[47] A similarly positioned residue in src SH3 also contains a productive nonnative conformation in the TSE.[48] Likewise, in nonnative regions of the cytochrome *b*562 intermediate, seven high ϕ values are observed ($0.4 \le \phi \le 1.0$).[46] Conversely, low ϕ values are found in regions of native-like structure in BPTI intermediates.[44]

In addition to Protein L, significant underreporting of the TSE's structural content by ϕ analysis also occurs with acyl phosphatase,[49–51] ubiquitin[27,28] and the B domain of Protein A[17] (Fig. 6). The ψ-determined TSEs for these proteins are extensive and contain persistent native-like tertiary interactions.

Unambiguous sites where ψ is unity indicate that the TSEs of acyl phosphatase and ubiquitin contain a four-stranded β sheet and an α helix. For these two α/β proteins, ψ analysis detects the presence of one and two additional long-range β strands than ϕ analysis identifies, respectively. We suspect that underreporting occurs with proteins having a TSE characterized by ϕ analysis as polarized, such as cold shock protein,[52] src SH3[53] and Protein G.[12]

Conclusion

We demonstrate that the highly studied TSE state of Protein L is extensive and has nonnative properties that likely arise due to the presence in the native state of backbone dihedral angles that are not highly populated at the earliest stage of folding. The TSE is significantly larger than the one identified by ϕ analysis, a result found in the three other proteins probed by ψ analysis (Fig. 6). The difference arises because ψ directly identifies interresidue contacts between known partners, while ϕ is native centric and the TSE can be less sensitive to energetic perturbations than the native state even for structured regions. These observations suggest that identification of a TSE as being diffuse, polarized or an expanded version of the native state based on ϕ analysis alone should be reconsidered.

Our results also emphasize that apparent agreement between the ϕ values and Gō-based models[43] can produce an overly optimistic view of these methods' ability to accurately determine TSE structures. Accurate modeling of protein folding remains an ongoing challenge, requiring the proper balancing of numerous factors in a changing contextual environment as the chain folds. ψ analysis should be applied to other proteins to address these outstanding issues, search for other nonnative TSE structures and provide a robust test set for benchmarking simulation, which may lead to better agreement between theory and experiment.

Materials and Methods

Folding measurements

The pseudo-wild-type sequence used has 64 amino acids, MEEVTIKANL IFANGSTQTA EFKGTFEKAT SEAYAYADTL KKDNGEWTVD VADKGYTLNI KFAG, which contains a Y47W mutation to enable fluorescence-monitored folding and unfolding. All variants are verified by DNA sequencing prior to expression. Purification uses either reverse-phase HPLC (C8 and C18 columns) or ion exchange (Amersham Biosciences Q Sepharose® Fast Flow) followed by gel-filtration HPLC (GE Healthcare HiPrep™ 16/60 Sephacryl™ S-100) in series. Protein samples are extracted in powder form following lyophilization. The purity of some mutants is verified by mass spectroscopy.

All measurements are taken in 100 mM NaCl and 50 mM Tris–HCl or Hepes (pH 7.5) buffer at 22 °C. Equilibrium folded and unfolded populations are measured via changes in circular dichroism using a Jasco 715 spectropolarimeter with a 1-cm path length. Kinetic data are collected using a Biologic SFM-400 stopped-flow apparatus and a PTI A101 arc lamp. Fluorescence spectroscopy uses $\lambda_{excite} = 285$ nm, and emission is observed at $\lambda = 310$–400 nm. Stock metal solutions of 250 mM $ZnCl_2$ and $NiCl_2$ are individually prepared in 10 mM HCl, as a concentrated source of metal cations, and are diluted to desired concentration prior to every experiment.

Data analysis

The kinetic data are analyzed using the "chevron analysis" of the denaturant dependence of folding rate constants[54] where ΔG_f, ΔG_u and ΔG_{eq} are linearly dependent on denaturant concentration:

$$\Delta G_{eq}([\text{den}]) = \Delta G^{H_2O} + m_0[\text{den}] \quad (4a)$$

$$\Delta G_f([\text{den}]) = RT\ln\left(k_f^{H_2O}\right) + m_f[\text{den}] \quad (4b)$$

$$\Delta G_u([\text{den}]) = RT\ln\left(k_u^{H_2O}\right) - m_u[\text{den}] \quad (4c)$$

where R is the universal gas constant and T is the temperature. The dependence on denaturant concentration [den], the m-values, reports on the degree of surface area burial during the folding process. The equilibrium values can be calculated from the kinetic measurements according to $\Delta G_{eq}([Me^{2+}]) = \Delta G_f([Me^{2+}]) - \Delta G_u([Me^{2+}])$ and $m_o = m_u + m_f$. To minimize extrapolation errors, we calculate ΔG_f and ΔG_u for strongly folding and unfolding conditions, respectively. ψ_0 values are determined from a simultaneous fit to the zero and high Me^{2+} chevrons, with ψ_0 being one of the fitting parameters and using Eq. (3). Parameters are fit using nonlinear least-squares algorithms implemented in the Microcal Origin software package.

All-atom MD simulations

Following similar protocols as in our previous studies,[55,56] we have simulated both amino and carboxy hairpins of Protein L with all-atom MD simulations, starting with nonnative random-coil conformations. The hairpin length is 24 and 21 aa for amino-terminal hairpin and carboxy hairpin, respectively. The hairpin fragments are solvated in a 54.5 Å × 48.0 Å × 45.0 Å water box, then sodium ions and chloride ions are added to neutralize and mimic the experimental environment (100 mM NaCl concentration). Both solvated systems contain approximately 12,000 atoms each. The IBM Blue Gene-optimized NAMD2[57] package is utilized for the MD simulations with the NPT ensemble at 1 atm and 295 K. The CHARMM (parameter set c32b1) force field is used for the protein,[58] and TIP3P water provides the model used as the explicit

solvent.[59,60] The PME (*P*article *M*esh *E*wald) method has been applied to treat the long-range electrostatic interactions and a 12-Å cutoff employed for the van der Waals interactions. The systems is first minimized and then followed by a 500,000-step equilibration. A snapshot taken during the equilibration is randomly selected as the starting point for the subsequent microseconds production runs. The time step for the production runs is 2 fs, and the SHAKE/RATTLE algorithm is applied.[61] The aggregate MD simulation time exceeds 4 μs.

Acknowledgements

We thank members of the Freed and Sosnick groups for helpful discussions and C. Antoniou for assistance in protein production. This work was supported, in part, by the National Institutes of Health Grant GM55694 (T.R.S.), National Science Foundation Grant CHE-1111918 (K.F.) and The University of Chicago-Argonne National Laboratory Seed Grant Program (T.R.S., Mike Wilde). R.H.Z. acknowledges the financial support from the IBM Blue Gene Program.

References

1. Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R., III & Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry*, **36**, 3373–3382.
2. Gu, H., Kim, D. & Baker, D. (1997). Contrasting roles for symmetrically disposed β-turns in the folding of a small protein. *J. Mol. Biol.* **274**, 588–596.
3. Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M. & Baker, D. (1998). The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* **284**, 807–815.
4. Kim, D. E., Fisher, C. & Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298**, 971–984.
5. Clementi, C., Garcia, A. E. & Onuchic, J. N. (2003). Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. *J. Mol. Biol.* **326**, 933–954.
6. Karanicolas, J. & Brooks, C. L., III (2002). The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **11**, 2351–2361.
7. Brown, S. & Head-Gordon, T. (2004). Intermediates and the folding of proteins L and G. *Protein Sci.* **13**, 958–970.
8. Yang, Q. & Sze, S. H. (2008). Predicting protein folding pathways at the mesoscopic level based on native interactions between secondary structure elements. *BMC Bioinformatics*, **9**, 320.
9. Zhao, L., Wang, J., Dou, X. & Cao, Z. (2009). Studying the unfolding process of protein G and protein L under physical property space. *BMC Bioinformatics*, **10**, S44.
10. Ejtehadi, M. R., Avall, S. P. & Plotkin, S. S. (2004). Three-body interactions improve the prediction of rate and mechanism in protein folding models. *Proc. Natl Acad. Sci. USA*, **101**, 15088–15093.
11. Koga, N. & Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* **313**, 171–180.
12. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of β-hairpin formation in protein G folding. *Nat. Struct. Biol.* **7**, 669–673.
13. Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8**, 602–605.
14. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y. & Baker, D. (2002). Accurate computer-based design of a new backbone conformation in the second turn of protein L. *J. Mol. Biol.* **315**, 471–477.
15. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
16. Plaxco, K. W., Simons, K. T., Ruczinski, I. & Baker, D. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*, **39**, 11177–11183.
17. Baxa, M., Freed, K. F. & Sosnick, T. R. (2008). Quantifying the structural requirements of the folding transition state of protein A and other systems. *J. Mol. Biol.* **381**, 1362–1381.
18. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using ψ-analysis. *J. Mol. Biol.* **337**, 463–475.
19. Pandit, A. D., Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2007). Characterizing protein folding transition states using Ψ-analysis. *Methods Mol. Biol.* **350**, 83–104.
20. Sosnick, T. R., Krantz, B. A., Dothager, R. S. & Baxa, M. (2006). Characterizing the protein folding transition state using ψ analysis. *Chem. Rev.* **106**, 1862–1876.
21. Sosnick, T. R. (2008). Kinetic barriers and the role of topology in protein and RNA folding. *Protein Sci.* **17**, 1308–1318.
22. Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T. & Sosnick, T. R. (2002). Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nat. Struct. Biol.* **9**, 458–463.
23. Krantz, B. A., Moran, L. B., Kentsis, A. & Sosnick, T. R. (2000). D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nat. Struct. Biol.* **7**, 62–71.
24. Krantz, B. A. & Sosnick, T. R. (2001). Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nat. Struct. Biol.* **8**, 1042–1047.
25. DeBartolo, J., Hocky, G., Wilde, M., Xu, J., Freed, K. F. & Sosnick, T. R. (2010). Protein structure prediction enhanced with evolutionary diversity: SPEED. *Protein Sci.* **19**, 520–534.
26. DeBartolo, J., Colubri, A., Jha, A. K., Fitzgerald, J. E., Freed, K. F. & Sosnick, T. R. (2009). Mimicking the folding pathway to improve homology-free protein structure prediction. *Proc. Natl Acad. Sci. USA*, **106**, 3734–3739.

27. Baxa, M. C., Freed, K. F. & Sosnick, T. R. (2009). ψ-Constrained simulations of protein folding transition states: implications for calculating ϕ values. *J. Mol. Biol.* **386**, 920–928.
28. Sosnick, T. R., Dothager, R. S. & Krantz, B. A. (2004). Differences in the folding transition state of ubiquitin indicated by ϕ and ψ analyses. *Proc. Natl Acad. Sci. USA*, **101**, 17377–17382.
29. Adhikari, A. N., Peng, J., Wilde, M., Xu, J., Freed, K. F. & Sosnick, T. R. (2012). Modeling large regions in proteins: applications to loops, termini, and folding. *Protein Sci.* **21**, 107–121.
30. Maity, H., Maity, M., Krishna, M. M., Mayne, L. & Englander, S. W. (2005). Protein folding: the stepwise assembly of foldon units. *Proc. Natl Acad. Sci. USA*, **102**, 4741–4746.
31. Haddadian, E. J., Gong, H., Jha, A. K., Yang, X., Debartolo, J., Hinshaw, J. R. *et al.* (2011). Automated real-space refinement of protein structures using a realistic backbone move set. *Biophys. J.* **101**, 899–909.
32. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L., Jr (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, **77**, 778–795.
33. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. (2005). Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl Acad. Sci. USA*, **102**, 13099–13104.
34. Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of α-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191.
35. Bai, Y., Zhou, H. & Zhou, Y. (2004). Critical nucleation size in the folding of small apparently two-state proteins. *Protein Sci.* **13**, 1173–1181.
36. Wallin, S. & Chan, H. S. (2006). Conformational entropic barriers in topology-dependent protein folding: perspectives from a simple native-centric polymer model. *J. Phys.: Condens. Matter*, **18**, S307–S328.
37. Ferguson, A., Liu, Z. & Chan, H. S. (2009). Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *J. Mol. Biol.* **389**, 619–636.
38. Paci, E., Lindorff-Larsen, K., Dobson, C. M., Karplus, M. & Vendruscolo, M. (2005). Transition state contact orders correlate with protein folding rates. *J. Mol. Biol.* **352**, 495–500.
39. Munoz, V. & Eaton, W. A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl Acad. Sci. USA*, **96**, 11311–11316.
40. Shoemaker, B. A., Wang, J. & Wolynes, P. G. (1999). Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. *J. Mol. Biol.* **287**, 675–694.
41. Alm, E. & Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl Acad. Sci. USA*, **96**, 11305–11310.
42. Galzitskaya, O. V. & Finkelstein, A. V. (1999). A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl Acad. Sci. USA*, **96**, 11299–11304.
43. Takada, S. (1999). Go-ing for the prediction of protein folding mechanisms. *Proc. Natl Acad. Sci. USA*, **96**, 11698–11700.
44. Bulaj, G. & Goldenberg, D. P. (2001). ϕ-Values for BPTI folding intermediates and implications for transition state analysis. *Nat. Struct. Biol.* **8**, 326–330.
45. Neudecker, P., Zarrine-Afsar, A., Choy, W. Y., Muhandiram, D. R., Davidson, A. R. & Kay, L. E. (2006). Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* **363**, 958–976.
46. Feng, H., Vu, N. D., Zhou, Z. & Bai, Y. (2004). Structural examination of ϕ-value analysis in protein folding. *Biochemistry*, **43**, 14325–14331.
47. Zarrine-Afsar, A., Dahesh, S. & Davidson, A. R. (2012). A residue in helical conformation in the native state adopts a β-strand conformation in the folding transition state despite its high and canonical ϕ-value. *Proteins*, **80**, 1343–1349.
48. Di Nardo, A. A., Korzhnev, D. M., Stogios, P. J., Zarrine-Afsar, A., Kay, L. E. & Davidson, A. R. (2004). Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc. Natl Acad. Sci. USA*, **101**, 7954–7959.
49. Pandit, A. D., Jha, A., Freed, K. F. & Sosnick, T. R. (2006). Small proteins fold through transition states with native-like topologies. *J. Mol. Biol.* **361**, 755–770.
50. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* **6**, 1005–1009.
51. Taddei, N., Chiti, F., Fiaschi, T., Bucciantini, M., Capanni, C., Stefani, M. *et al.* (2000). Stabilisation of α-helices by site-directed mutagenesis reveals the importance of secondary structure in the transition state for acylphosphatase folding. *J. Mol. Biol.* **300**, 633–647.
52. Garcia-Mira, M. M., Boehringer, D. & Schmid, F. X. (2004). The folding transition state of the cold shock protein is strongly polarized. *J. Mol. Biol.* **339**, 555–569.
53. Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.* **5**, 714–720.
54. Matthews, C. R. (1987). Effects of point mutations on the folding of globular proteins. *Methods Enzymol.* **154**, 498–511.
55. Das, P., King, J. A. & Zhou, R. (2011). Aggregation of γ-crystallins associated with human cataracts via domain swapping at the C-terminal β-strands. *Proc. Natl Acad. Sci. USA*, **108**, 10514–10519.
56. Zhou, R., Berne, B. J. & Germain, R. (2001). The free energy landscape for β hairpin folding in explicit water. *Proc. Natl Acad. Sci. USA*, **98**, 14931–14936.
57. Kumar, S., Huang, C., Zheng, G., Bohm, E., Bhatele, A., Phillips, J. C. *et al.* (2008). Scalable molecular dynamics with NAMD on the IBM Blue Gene/L system. *IBM J. Res. Dev.* **52**, 177–188.
58. Brooks, B. R., Brooks, C. L., III, Mackerell, A. D., Jr, Nilsson, L., Petrella, R. J., Roux, B. *et al.* (2009).

CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614.
59. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Comput. Chem.* **79**, 926–935.
60. Neria, E. & Karplus, M. (1996). A position dependent friction model for solution reactions in the high friction regime: proton transfer in triosephosphate isomerase (TIM). *J. Chem. Phys.* **105**, 10812–10818.
61. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341.
62. Sato, S. & Fersht, A. R. (2007). Searching for multiple folding pathways of a nearly symmetrical protein: temperature dependent ϕ-value analysis of the B domain of protein A. *J. Mol. Biol.* **372**, 254–267.