Intramolecular Cross-Linking Evaluated as a Structural Probe of the Protein Folding Transition State†

Ali T. Shandiz, Benjamin R. Capraro, and Tobin R. Sosnick*

Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637

Received May 29, 2007; Revised Manuscript Received September 4, 2007

ABSTRACT: We examine the utility of intramolecular covalent cross-linking to identify the structure present in the folding transition state. In mammalian ubiquitin, cysteine residues located across two β-strands are cross-linked with dichloroacetone. The kinetic effects of these covalent cross-links in ubiquitin, and engineered disulfide bonds in src SH3 (Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97,* 7084−7089), are compared to the results of ψ-analysis where strand association is stabilized by metal ion binding to engineered bihistidine sites (Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004) *J. Mol. Biol. 337,* 463−75) at the same positions. The results for the two methods agree at some of the sites. The cross-linking $\phi^{\text{crosslink}}$-values agree with their corresponding ψ-values when they have both have values of zero or one, which represent the absence and presence of native structure, respectively. When $\phi^{\text{crosslink}} > \psi$, the apparent inconsistency is rationalized by the difference between each method's mode of stabilization; cross-linking reduces the configurational entropy of the unfolded state whereas metal binding directly stabilizes the native state. However, when the cross-linking ϕ-values are smaller than their corresponding ψ-values, the apparent underestimation of structure formation is difficult to rationalize while retaining the assumption that the cross-link exclusively affects the entropy of the unfolded state. The interpretation also is problematic for data on cross-links located across strands which are not hairpins, and hence, these sites are likely to be of limited utility in folding studies. We conclude that cross-linking data for sites on hairpins generally report on the amount of structure formed within the enclosed loop while the metal binding data report on the amount structure formed at the site itself.

A continuing goal of experimental protein folding studies is to elucidate the structure of the transition state ensemble (TSE[1]) in the hopes of understanding the critical steps of the folding process. To identify regions structurally important in the TSE, many studies observe the changes in kinetic response that result from a perturbation. In mutational ϕ-analysis, the perturbation takes the form of point mutations engineered throughout the protein (*1*−*3*). The accompanying changes in folding rates report on the mutated side-chain's level of energetic involvement in the TS. The extent of involvement is quantified as the ϕ-value, defined as the change in folding activation energy relative to the change in equilibrium stability, $\phi = \Delta\Delta G^{\ddagger}/\Delta\Delta G_{\text{eq}}$.

A new method, ψ-analysis, employs metal ion binding to engineered bihistidine (biHis) sites as the means of perturbation (*4*−*9*) (Figure 1). An increase in divalent metal ion concentration stabilizes the interaction between the two histidine partners. Analogous to mutational ϕ-analysis, the relative amount of binding energy is quantified as $\psi = \partial\Delta\Delta G^{\ddagger}/\partial\Delta\Delta G_{\text{eq}}$. The change in kinetic response due to metal binding reflects the difference between the binding affinity of the site in the TSE relative to the affinity in the unfolded state. A region that is native-like in the TS will realize its full metal binding energy in the TS ($\psi = 1$), while a region that is partially or unstructured in the TS will not ($0 \leq \psi < 1$). By bridging the side-chains of two residues distant in sequence but structurally local, metal ion binding serves as a direct probe of long-range interactions and, hence, directly reports on chain topology in the TSE.

Another method designed for exploring long-range interactions and chain topology utilizes covalent cross-links. Engineered disulfide bonds created using bicysteine (biCys) substitutions have been utilized in kinetic folding studies (*10*−*12*). Using the unlinked biCys variant as the wild-type condition, the change in folding activation energy relative to the change in equilibrium stability due to the formation of the cross-link is calculated as a $\phi^{\text{crosslink}}$-value, similar to mutational ϕ-analysis. Unlike either mutagenesis or metal ion binding, cross-linking is believed to act primarily on the unfolded state, stabilizing the protein by reducing the configurational entropy of the unfolded chain. This change in entropy is referred to as "loop closure" entropy (*13*). The $\phi^{\text{crosslink}}$-value reflects the degree to which the loop enclosed by the cross-link has lost configurational entropy in the TS.

† This work is supported by a research (GM55694) and a training grant (GM007183-32) from the NIH. B.R.C. was supported by the University of Chicago NIH Undergraduate Training Program in Physical and Chemical Biology.

* To whom correspondence should be addressed: trsosnic@uchicago.edu, tel 773-218-5950, fax 702-0439.

[1] Abbreviations: biCys, bicysteine; biHis, bihistidine; DCA, dichloroacetone; GdmCl, guanidinium chloride; TSE, transition state ensemble; Ub, mammalian ubiquitin; ψ_o, fraction of the native metal ion binding energy states realized in the transition state, extrapolated to zero metal ion concentration.

10.1021/bi701042e CCC: $37.00
Published on Web 11/07/2007

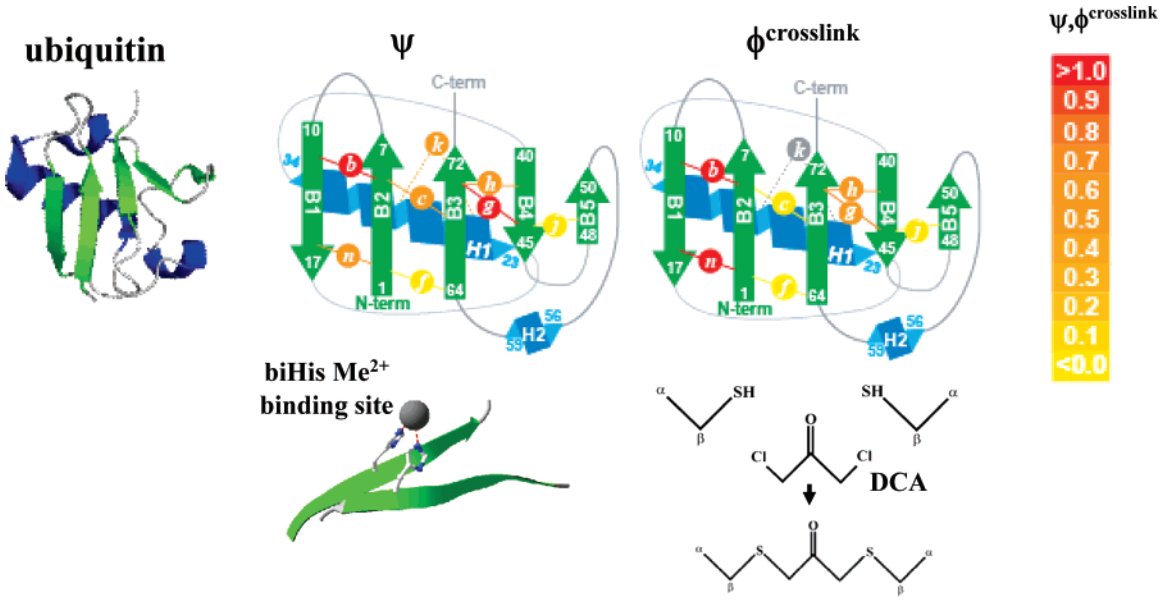

FIGURE 1: Comparison of metal ion binding and cross-linking results. Ub structure (*27*) and schematic representations of ψ-analysis and cross-linking results. Below the diagrams are illustrations of a biHis binding site and the chemical constituents involved in DCA cross-linking at a biCys site. The biHis/biCys sites are shown as circles with italic letters; each site was studied individually. The color intensity represents the value of ψ- or ϕ-value at each site.

The amount of entropy lost is an indication of the difference between the extents to which the ends of the loop are closed in the TS before and after the introduction of the cross-link. If the loop is unstructured or completely closed in the TS, the change in kinetic response is expected generate a $\phi^{crosslink}$-value of zero or one, respectively (Figure 2A,B).

The interpretations of the results of cross-linking experiments are sometimes at odds with those of the results of mutational ϕ-analysis. In experiments performed in barnase by Fersht and co-workers, the results of the two methods present conflicting pictures of the TS (*10*). As expected, a cross-link in a region believed to be unstructured in the TS results only in changes to the rate of unfolding. Contrary to expectation, however, a second cross-link in a region believed to be structured causes changes in both the rate of folding and unfolding. The result of the second cross-link suggests that a region believed to be structured in the TS based on mutational ϕ-analysis is not fully structured. Cross-linking experiments by Baker and co-workers on src SH3 also resulted in some ambiguities (*11*). Of the three sites they examined, only the results for one site readily conformed to expectations based on mutational work. The results of cross-linking work in CD2 by Clark and co-workers were similarly mixed (*12*). It is unclear whether the discrepancies between cross-linking and mutational ϕ-analysis are due to interpretation of the cross-linking or the mutational data. Kiefhaber and co-workers removed two naturally occurring disulfide bonds in tendamistat and concluded that a local hairpin was formed (*14*) although no direct comparison with ϕ-values was performed.

Here we investigate the suitability of cross-links as probes of long-range interactions in the TS. We also consider the utility of the artificial cross-linker dichloroacetone (DCA) for intramolecular cross-linking. We find that the cross-linking results for Ub cannot always be interpreted in a manner consistent with both the picture of the TSE presented by ψ-analysis and the prevailing model of protein stabilization by a cross-link. Our results suggest that the sites at which the interpretations of the results of the two methods are consistent are the sites that span topologically simple loops (e.g., a hairpin). Sites at which the results of the two methods cannot be reconciled appear to span more topologically complicated regions.

In addition, the apparent discrepancy likely is due to the different manner in which each method stabilizes the protein. Ideally, cross-linking stabilizes the protein by reducing the configurational entropy of the unfolded chain whereas metal-ion binding stabilizes native-like structures where the biHis sites are binding competent. Due to a variety of factors, the effects of cross-linking may not always be restricted to the unfolded chain. When cross-linking affects both the denatured and native states, interpretation becomes difficult. Cross-linking across topologically complicated regions may increase the likelihood that the effect will not just be on the unfolded chain. We conclude that cross-linking can be an effective tool for identifying structure in the TS if appropriately placed and interpreted according to the amount of structure formed within the enclosed loop, rather than at the site of the cross-link itself.

MATERIALS AND METHODS

Expression and Purification of Ubiquitin. The biCys double mutants of Ub F45W were created sequentially using Stratagene's QuikChange mutagenesis kit. An additional H68N substitution was introduced into all variants for the purpose of remaining consistent with the previously produced biHis variants used for metal ion binding experiments. Two variants (sites *c* and *n*) were introduced into a previously engineered background containing additional R42H and V70H substitutions. All proteins were expressed and purified as previously described(*15*).

Expression and Purification for the src SH3 Domain. For the src SH3 domain, the nucleotide sequence for residues T81−D141 of c-Src kinase (PDB accession 1FMK (*16*)) was obtained from a synthetic gene, introducing a two-residue

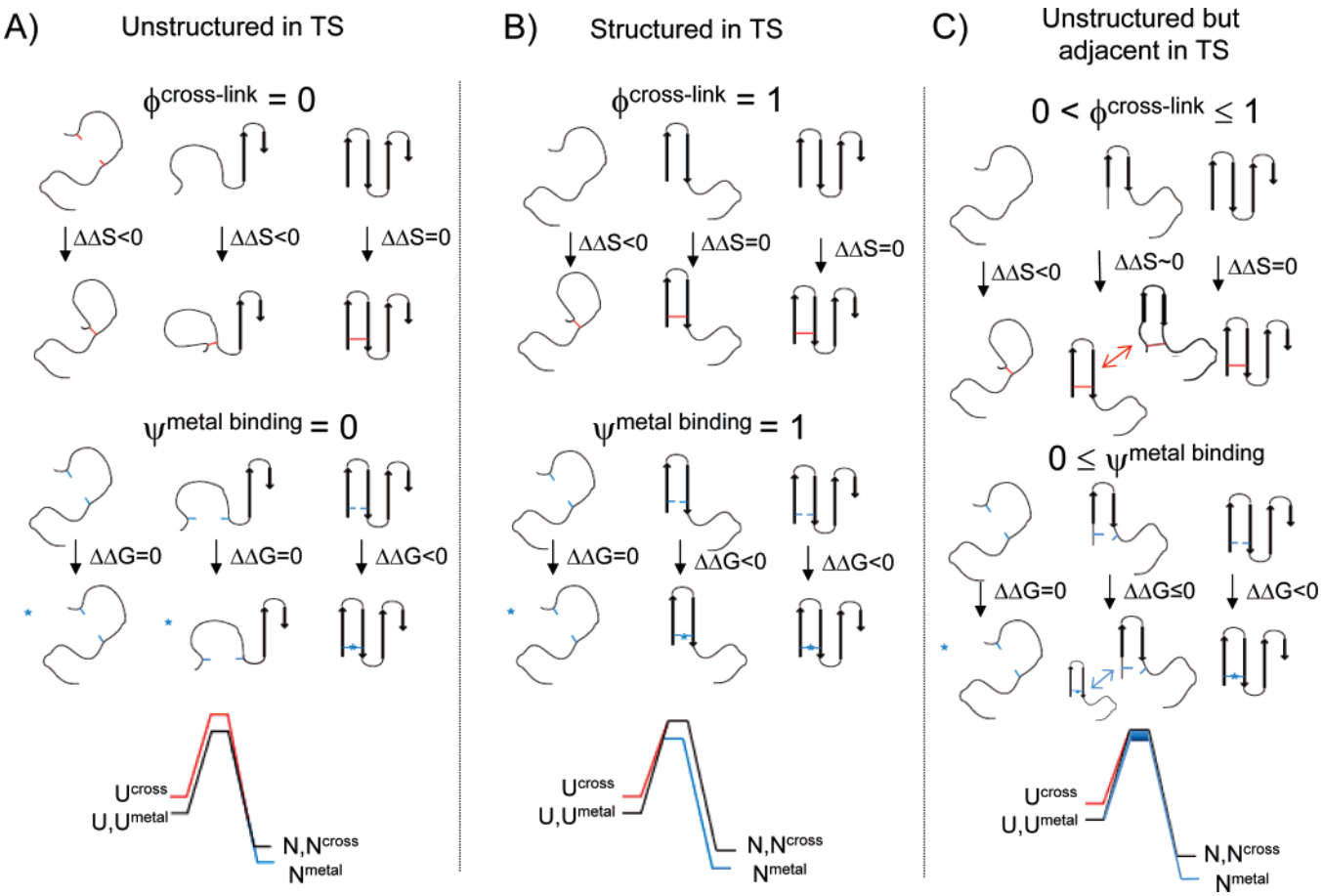

FIGURE 2: Interpretation of cross-linking and metal ion binding data. (A) In this example, the region encompassing the cross-link is unstructured and the biHis binding site is absent in the TS. Although the mode of stabilization is different for the two methods, they should produce similar outcomes. The energy diagram (bottom) illustrates the effect on the relative free energies of U, TS, and N with no cross-link (black line) and after cross-linking (red line) or in the presence of divalent metal ions (blue line). The entropic model assumes the effect of the cross-linker (red lines) to be solely due to the loss of configurational entropy in the unfolded structure. When the structure at the cross-linked site is not formed in TS, the losses of entropy of U and TS are altered equally, thereby keeping the relative barrier to folding unchanged and $\phi^{\text{crosslink}} = 0$. In contrast, metal ion binding acts only on a native-like structure. When the biHis site (blue lines) is not formed in the TS, metal ion binding only stabilizes N, increasing the relative height of the barrier to unfolding to produce a value of $\psi = 0$. (B) In the situation where the region enclosed by the cross-link is structured and the biHis site formed in the TS, the two methods also should produce similar outcomes. Cross-linking at a structured region results in no additional loss of entropy in the TS. Hence, only the entropy of the U state is affected. The barrier to folding is reduced and a value of $\phi^{\text{crosslink}} = 1$ is generated. With the native-like biHis site in the TS, metal ion binding stabilizes both the TS and N, and only reduces the relative height of the folding barrier to produce $\psi = 1$. (C) When the biHis site is not formed and the cross-link is located across an unstructured region, but adjacent to native-like structure in the TS, the two methods may produce different values with $\phi^{\text{crosslink}} > \psi$. Due to the structure present within the enclosed loop, the loss of entropy due to loop-closure is smaller in the TS than in the U state. The barrier to folding is lowered and $\phi^{\text{crosslink}}$ can be fraction with a value approaching unity depending on the location of the structure. Alternatively, the introduction of the cross-link can induce structure formation up to the cross-link itself, resulting in $\phi^{\text{crosslink}} \sim 1$ (red arrow). The outcome of metal binding is similarly variable. Without the binding site formed in the TS, metal binding initially stabilizes only the native state, leading to a $\psi^{\text{metal binding}} \sim 0$. With an increasing degree of metal-induced stabilization, the equilibrium in the TS can shift toward the site being formed (blue arrow), leading to a ψ-value greater than zero.

Gly-Ser N-terminal tail. BiHis variants for metal binding studies, and the pseudo wild-type background H46N, were obtained using QuikChange mutagenesis; residues are numbered as per PDB accession 1SRL (*17*), in accord with previous folding studies (*18*). Proteins were expressed in *Escherichia coli*, extracted from inclusion bodies, and purified by anion exchange chromatography and further by reverse-phase HPLC. Masses were verified by mass spectrometry, and proteins were stored lyophilized. Samples for site *b* (E17H/S25H/H46N) were found to contain a minor population of variants with the unnatural methionine introduced for protein expression uncleaved (20% relative ionization), and were used without further purification.

Dichloroacetone Cross-Linking. The biCys mutants of Ub were cross-linked by the bifunctional thiol-specific reagent DCA using a modified version of a protocol presented elsewhere (*19*). The reaction was performed on ice to eliminate the occurrence of side-reactions, and all components were prechilled prior to the initiation of the reaction. Reaction stocks of each Ub variant to be cross-linked were made up in 200 mM triethanolamine buffer, pH 5.5, from high concentration unbuffered guanidinium chloride (GdmCl) stocks of the protein. These stocks were diluted to a final concentration of 150 μM with a 0.4× volume of ice-cold 250 mM sodium borate, pH 8.3. The reaction was initiated by the addition of four molar equivalents of 100 mM DCA in dimethylformamide and allowed to proceed for 30 min. The reaction was quenched by the addition of 50 mM β-mercaptoethanol. After 30 min, the pH of the reaction mixture was dropped to 5.0 with concentrated acetic acid

and the cross-linked species was isolated by reverse-phase HPLC. Successful cross-linking was confirmed by electrospray TOF-mass spectrometry by a change in mass of 54 Da.

Kinetic Experiments. Kinetic experiments were performed, as previously described, using a Biologic stopped-flow apparatus (*15*). Experiments were conducted at 20 °C, 50 mM Hepes (pH 7.6), and a final protein concentration of 1.5 μM. For experiments performed with unmodified variants, 1 mM of the reductant dithiothreitol was added to all buffers to prevent disulfide bond formation. Native fluorescence from Trp45 (excitation wavelength 280 ± 10 nm; emission wavelength (bandpass filter, 350 ± 50 nm) was used to monitor the folding and unfolding reactions.

The kinetic data were analyzed using chevron analysis with the free energy of equilibrium folding and the activation free energy for kinetic folding and unfolding, ΔG_{eq}, $\Delta G^{\ddagger}_{f}$, and $\Delta G^{\ddagger}_{u}$, assumed to be linearly dependent on denaturant concentration:

$$\Delta G_{eq}([\text{GdmCl}]) = \Delta G_{\text{H2O}} - m^{\circ}[\text{GdmCl}]$$

$$\Delta G^{\ddagger}_{f}([\text{GdmCl}]) = RT \ln k_{f}^{\text{H2O}} - m_{f}[\text{GdmCl}]$$

$$\Delta G^{\ddagger}_{u}([\text{GdmCl}]) = RT \ln k_{u}^{\text{H2O}} - m_{u}[\text{GdmCl}]$$

The slopes m°, m_{f}, and m_{u} represent the difference in the amount of denaturant sensitive surface area buried between the initial and final states for the transition under consideration. Parameters were fit using the nonlinear least-squares algorithm available in the Microcal Origin software package.

Experiments with SH3 were performed in 50 mM HEPES, pH 7.5, 10 °C, at a final protein concentration of 5 μM. Experiments with metal used 1 mM $CoCl_2$, and in the absence of metal, 1 mM EDTA.

RESULTS

The interpretation of the cross-linking data is evaluated in light of the comprehensive picture of the TSE of mammalian ubiquitin (Ub) produced by ψ-analysis (*4*, *9*). Ubiquitin's TSE consists of a topologically native-like core composed of the carboxy-terminus of the α-helix docked against four partially formed strands of the β-sheet network (Figure 1). Several of the sites on the periphery of this core have fractional ψ-values. A fractional ψ-value can be interpreted as a biHis site that either is native-like in a fraction of the TSE at the level given by the ψ-value or has weaker metal ion binding affinity in the TSE, or some combination thereof (*9*, *20*). Nonetheless, the five biHis sites with ψ-values of unity generate an unambiguous picture of the TS kernel.

Cross-Linking. Seven individual biCys variants of ubiquitin are produced (Figure 1). These sites correspond to a subset of the engineered biHis metal-binding sites previously used in the ψ-analysis study (*4*). The cross-link is created by substituting a pair of cysteines for the two histidines, followed by the addition of a small molecule, dichloroacetone (DCA), to create a $C_{\alpha}-S-CH_2-(C{=}O)-CH_2-S-C_{\alpha}$ linkage. Six sites are introduced across adjacent strands spanning the β-sheet network while one site is located at the amino-terminus of the α-helix. All but two of the cross-linked variants are in the background of the pseudo-WT (F45W H68N); sites *c* and *n* contained additional histidine mutations at R42 and V70. These additional histidine mutations, generated to produce a biHis/biCys combination for a complementary study, had no effect on the results of the cross-linking as determined by comparison of cross-linking at other sites in both pseudo-WT and biHis backgrounds (data not shown).

Cross-linking with DCA is possible at locations similar to those previously tested for engineered disulfides in other proteins (*10*−*12*). For all the β-sheet sites, the DCA cross-link added the expected 54 Da shift in the molecular weight according to electrospray mass spectrometry (data not shown). However, the α-helix site could not be cross-linked. The proximity of the cysteines at the α-helix site (*i*, *i* + 4) may cause unfavorable side-chain geometries or the helical backbone conformation may be intolerant of strain introduced by the cross-linker.

The cross-linked proteins exhibited varying degrees of enhanced stability (Table 1). However, none of the cross-linked proteins experienced stability gains fully consistent with those predicted by the Flory and Jacobson−Stockmayer theory for Gaussian chains, wherein the entropic cost of loop closure of *n* segments is $\Delta S = (-3/2)R \ln(n\pi/2)$ (*13*, *21*, *22*). Baker and co-workers also observed discrepancies in src SH3 (*18*) while Bowler and co-workers observed discrepancies in their cytochrome *c* studies (*23*).

Kinetic Analysis. The cross-linking results have marginal agreement with the metal ion binding experiments. Of the six sites successfully cross-linked, only three sites produce the same results as ψ-analysis (Figures 1, 3, Table 1). These three sites, sites *b*, *j*, and *h*, are located across the protein and exhibit the three general classes of results, unity, zero, and intermediate $\phi^{\text{crosslink}}$-values. Site *b* is formed across the β1-β2 hairpin. This site's $\phi^{\text{crosslink}}$-value of 1.07 ± 0.03 indicates that the loop is closed in the TS of the protein lacking the DCA cross-link. This result is consistent with the ψ-analysis data which indicated that the hairpin is native-like in the TS; the ψ-values for site *b* and the adjacent biHis site *a* were 1.06 ± 0.01 and 1.32 ± 0.13, respectively. This outcome is illustrated in Figure 2B.

For site *j*, located across the β4-β5 hairpin, the $\phi^{\text{crosslink}}$-value is very near zero (0.12 ± 0.09). This result indicates that the cross-link joins parts of the chain that do not associate in the TS of the protein lacking the cross-link. The metal ion binding data for site *j* agrees with this interpretation, having a ψ-value also close to zero (0.02 ± 0.00). This outcome is illustrated in Figure 2A.

Finally, site *h*, across strands β3 and β4 (residues 42−70), has a fractional $\phi^{\text{crosslink}}$-value, 0.42 ± 0.04 (Figure 2). The ψ-value at this site is in good agreement (ψ = 0.57 ± 0.07), although it may be coincidental (see Discussion). The $\phi^{\text{crosslink}}$-values of the three remaining sites significantly differ from their corresponding ψ-values. In two cases, the $\phi^{\text{crosslink}}$-values are lower than expected, giving the appearance that these sites are less structured in the TS than indicated by their corresponding ψ-values. One of these two positions, site *c*, is between strands β2 and β3 (residues 6−68). Its $\phi^{\text{crosslink}}$-value is nearly zero (0.12 ± 0.09), despite the corresponding biHis site having a substantial ψ-value (0.52 ± 0.02). Site *g*, which cross-links strands β3 and β4 (residues 44−70), has a fractional $\phi^{\text{crosslink}}$-value of 0.32 ± 0.01, considerably lower than the ψ-value of 1.25 ± 0.04. If site

Table 1: Equilibrium and Kinetic Parameters for DCA Cross-Linking in Ub[a]

site	mutation[b]	$\ln(k_f)$	$\ln(k_u)$	m_f	m_u	m_o	$\Delta\Delta G_f^{\ddagger}$	$\Delta\Delta G_{eq}$	$\Delta\Delta G_{predict}$[b]	$\phi^{crosslink}$	ψ^c
b	K6C-T12C	0.28 ± 0.03	1.44 ± 0.03	−1.47 ± 0.03	0.72 ± 0.01	−2.19 ± 0.03	−1.18 ± 0.03	−1.18 ± 0.04	−2.09	1.07 ± 0.03	1.06 ± 0.01
	cross-linked	2.31 ± 0.04	1.56 ± 0.05	−1.51 ± 0.04	0.74 ± 0.03	−2.25 ± 0.04					
c	K6C-H68C	0.39 ± 0.05	2.07 ± 0.04	−1.39 ± 0.06	0.71 ± 0.02	−2.10 ± 0.05	−0.09 ± 0.08	−0.75 ± 0.10	−4.01	0.12 ± 0.09	0.52 ± 0.03
	cross-linked	0.54 ± 0.12	0.94 ± 0.10	−1.33 ± 0.15	0.52 ± 0.05	−1.85 ± 0.13					
g	I44C-V70C	1.25 ± 0.04	1.64 ± 0.06	−1.50 ± 0.05	0.63 ± 0.05	−2.13 ± 0.05	−0.86 ± 0.03	−2.71 ± 0.12	−3.27	0.32 ± 0.01	1.25 ± 0.04
	cross-linked	2.73 ± 0.04	−1.53 ± 0.18	−0.77 ± 0.05	0.99 ± 0.10	−1.76 ± 0.05					
h	R42C-V70C	1.77 ± 0.04	1.07 ± 0.03	−1.60 ± 0.04	0.71 ± 0.02	−2.31 ± 0.04	−0.49 ± 0.06	−1.17 ± 0.07	−3.34	0.42 ± 0.04	0.57 ± 0.07
	cross-linked	2.61 ± 0.11	−0.10 ± 0.06	−1.12 ± 0.05	0.96 ± 0.05	−2.09 ± 0.05					
j	I44C-Q49C	1.19 ± 0.06	1.56 ± 0.06	−1.42 ± 0.08	0.68 ± 0.03	−2.10 ± 0.08	−0.11 ± 0.07	−0.84 ± 0.09	−1.96	0.13 ± 0.07	0.02 ± 0.00
	cross-linked	1.38 ± 0.10	0.29 ± 0.09	−1.51 ± 0.11	0.54 ± 0.05	−2.05 ± 0.10					
n	Q2C-E16C	0.57 ± 0.08	1.90 ± 0.09	−1.31 ± 0.09	0.75 ± 0.05	−2.06 ± 0.08	−0.96 ± 0.06	−0.96 ± 0.09	−2.76	0.99 ± 0.07	0.53 ± 0.06
	cross-linked	2.21 ± 0.07	1.88 ± 0.07	−1.50 ± 0.06	0.70 ± 0.05	−2.20 ± 0.06					

[a] To minimize extrapolation errors, values for $\Delta\Delta G_f^{\ddagger}$, $\Delta\Delta G_{eq}$, and $\phi^{crosslink}$ are calculated using the values determined at 2 and 6 M GdmCl, respectively, generated from a simultaneous fit to the two relevant chevrons, with the parameter of interest being one of the fitting parameters. Units are kcal mol^{-1} (free energies) or kcal mol^{-1} M^{-1} (*m*-values). All mutants were produced in pseudo-WT background of F45W H68N, except site *c* (R42H, F45W, V70H) and site *n* (R42H, F45W, H68N, V70H). [b] $\Delta\Delta G_{predict} = (-3/2)\,RT\ln(n\pi/2)$ for *n* segments (*13*, *21*, *22*). [c] ψ-values listed are the values extrapolated to zero metal concentration and corrected for the insertion of the two histidines ("ψ_o^{corr}" in ref (*4*)).

FIGURE 3: Kinetic analysis of DCA cross-linking. The denaturant dependence of folding for the cross-linked (red circles) and the unmodified biCys variants (black circles) for the six Ub cross-linked sites is shown. The subscript of the ϕ-value indicates the cysteine residues involved. Curves are the fits to the chevron data (Table 1).

g is structured in the TS, as supported by the ψ-value, a $\phi^{crosslink}$ near unity would be expected for this site.

The third position, site *n* located at the end of the $\beta1$-$\beta2$ hairpin, has a $\phi^{crosslink}$ of unity (0.99 ± 0.07). The corresponding ψ-value is significantly lower (0.53 ± 0.06). Although there appears to be a discrepancy between the two methods, in fact the outcome can be rationalized. Because the portion of the hairpin closer to its turn already is formed in the TS ($\psi \sim 1$ at sites *a* and *b*), most of the chain's configurational entropy is already lost in the TS. Hence, the addition of the cross-link has much less effect on the entropy of the TS than on the U-state and a $\phi^{crosslink} \sim 1$ is expected even though the ends are non-native. This outcome where $\phi^{crosslink} > \psi$ is illustrated in Figure 2C.

ψ-Analysis on SH3. To provide a direct comparison between cross-linking and ψ-analysis data on another protein and to test our interpretation of situations where $\phi^{crosslink} > \psi$-value, ψ-analysis is applied to two sites on the RT loop of the all-β protein src SH3 (Figure 4, Table 2). For an engineered disulfide bond across the far end of the RT loop (between residues 9 and 33), Baker and co-workers measured a $\phi^{crosslink}$ value of 0.45 (*18*). We introduce a biHis site at a nearby position, residues 11 and 31. The addition of 1 mM $CoCl_2$ increases the protein's stability by 1.15 ± 0.03 kcal mol^{-1}, while the folding activation free energy barely changes (the 0.13 ± 0.04 kcal mol^{-1} increase in folding activation free energy may be caused by ion binding in the denatured state or non-native interactions, see (*24*, *25*)). Nevertheless, the resulting ψ-value is near zero, indicating that this site becomes structured only after the rate-limiting step in folding.

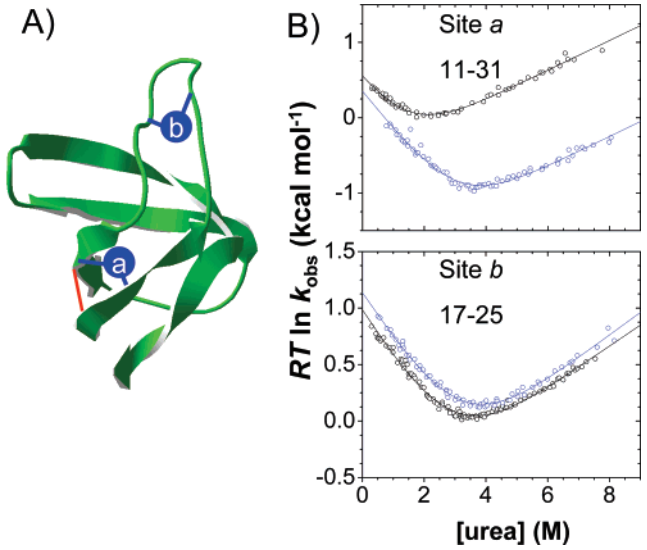

FIGURE 4: ψ-Analysis on the RT loop of the src SH3 domain. (A) SH3 structure (*16*) with prior site of disulfide cross-linking where $\phi^{crosslink} = 0.45$ (*18*) (red line; residues 9−33) and biHis sites employed in this study (blue lines): site *a*, residues 11−31; site *b*, residues 17−25. (B) Denaturant dependence of folding rates in the presence of 1 mM EDTA (black points), and in the presence of 1 mM $CoCl_2$ (gray points). Curves are simultaneous chevron fits to the 0 and high metal datasets, yielding the kinetic and thermodynamic parameters summarized in Table 2. The outcome with a near-zero ψ-value and fractional $\phi^{crosslink}$-value is presented in Figure 2C.

Another biHis site is inserted at residues 17 and 25, a site closer to the turn in the RT loop. Here, metal binding primarily affects the folding rate. Hence, residues 17 and 25 are forming a biHis site in a significant fraction of the TSE. This result, in conjunction with the small relative change in folding rates for the biHis site at residues 11 and 31, rationalizes the fractional $\phi^{crosslink}$ observed at residues 9 and 33. Although the ends of loop are largely unstructured, some amount of structure is formed interior to the disulfide bond. This structure reduces the entropy of the enclosed chain so that the stabilization imparted by the disulfide cross-link is diminished. As a result, a fractional $\phi^{crosslink}$ is expected, consistent with the observed value of 0.45. This situation is similar to the results for Ub site *n*, and the relevant energy diagrams are depicted in Figure 2C.

DISCUSSION

Cross-links have been used as a probe of long-range interactions in the TS, but some results have proven problematic for interpretation (*10*−*12*). Our interpretation of the DCA cross-linking data for Ub benefits from the extensive characterization of Ub's TSE by ψ-analysis, as well as the results of previous cross-linking studies on other proteins including src SH3 where we have applied ψ-analysis. By examining sites where both cross-linking and ψ-analysis have been applied, we have more information with which to interpret the cross-linking results.

There is little reason *a priori* to expect that the results of cross-linking will disagree with the results of metal binding. However, a comparison highlights the apparent inconsistencies between the results of the two methods (Figures 1 and 2). In some locations, cross-linking suggests more structure is formed in the TS than metal binding does while at other locations, the reverse is observed. It is difficult to account for these differences with a single explanation, because one method does not consistently report more structure than the other.

Nevertheless, the comparison to the results of previous cross-linking studies allows us to generalize the common structural or topological features characteristic of certain outcomes (Figure 2). Our interpretation is based on the premise that the cross-linking data report on the structure formed within the enclosed loop while the metal binding data report on the amount of structure formed at the biHis site itself. The cross-link is assumed to exclusively affect chain conformational entropy and have minimal effect on native-like structures. In contrast, metal ion binding generally acts only on native-like structures, stabilizing them once the biHis site is formed.

Although the mechanism is different, the interpretation is likely to be the same when both the $\phi^{crosslink}$ and the ψ-values are either unity or zero (Figure 2A,B). When the $\phi^{crosslink}$-value is unity, the cross-link joins segments of the chain that already are associated in the TS of the unlinked version of the protein. This interpretation is consistent with a native-like biHis site that is formed in the entire TS ensemble, i.e., a ψ-value of unity. In addition, a unity $\phi^{crosslink}$-value could also be observed at a site that is formed in the entire TS ensemble but which has a fractional ψ-value due to a weaker metal ion binding affinity in TS than in the native state. When the $\phi^{crosslink}$ -value is zero, the cross-link brings together the ends of a loop that are unstructured in the TS of the unmodified protein. Again, this interpretation is consistent with a biHis site that is absent in the entire TS ensemble and has a ψ-value of zero.

Cross-Linking and Topology. We believe that the outcome of a cross-linking experiment is dependent upon the topology of the region enclosed by a cross-link and the number of secondary structural elements enclosed in the loop. The results for the three sites that span local β-hairpins were the most consistent with both the ψ-analysis results and the assumption that cross-links stabilize proteins by exclusively affecting the unfolded states entropy. The topology of the regions enclosed by the cross-links at these sites corresponds to the "single loop topology" described by Thornton for many naturally occurring disulfides (*26*), wherein the cross-link closes off a native loop consisting of a single 180° chain reversal and no more than two secondary structural elements. For two of these "single loop" sites, sites *b* and *j*, the $\phi^{crosslink}$-values of nearly 1 and 0 are essentially identical to their respective ψ-values of 1.06 ± 0.01 and 0.02 ± 0.00. As discussed above, these $\phi^{crosslink}$-values and ψ-values can be interpreted identically. The $\phi^{crosslink}$-value and ψ-value of unity observed for site *b* can both be interpreted as evidence that the β1-β2 hairpin is formed in the TS. The $\phi^{crosslink}$-value of zero observed for the cross-link at site *j* indicates that the loop formed by strands β4 and β5 is unstructured in the TS. This result is consistent with the two near-zero ψ-values on this small hairpin.

Along with site *b*, site *n* spans strands β1 and β2. The near unity $\phi^{crosslink}$-value appears inconsistent with the fractional ψ-value observed for this site (0.53 ± 0.06). However, the two results can be rationalized after accounting for the two methods' different modes of stabilization. The fractional ψ-value may be interpreted as either a biHis site

Table 2: Kinetic Parameters for Divalent Ion Binding in src SH3[a]

site	mutation	$RT\ln(k_f)$ (0 M)	$RT\ln(k_u)$ (5 M)	m_f	m_u	$\Delta\Delta G_f^{\ddagger}$	$\Delta\Delta G_{eq}$
a	V11H-R31H	0.47 ± 0.03	0.43 ± 0.01	-0.54 ± 0.05	0.20 ± 0.01	0.13 ± 0.04	-1.15 ± 0.03
	1 mM Co^{2+}	0.34 ± 0.07	-0.85 ± 0.09	-0.51 ± 0.02	0.20 ± 0.01		
b	E17H-S25H	0.96 ± 0.01	0.098 ± 0.01	-0.44 ± 0.01	0.19 ± 0.01	-0.15 ± 0.02	-0.10 ± 0.02
	1 mM Co^{2+}	1.12 ± 0.03	0.15 ± 0.04	-0.41 ± 0.01	0.20 ± 0.01		

[a] To minimize extrapolation errors, values for $\Delta\Delta G_f^{\ddagger}$ and $\Delta\Delta G_{eq}$ upon the addition of 1 mM Co^{2+} are obtained from the change in ln k_f and ln k_u at 0 and 5 M urea, generated from a simultaneous fit to the two relevant chevrons, with the parameter of interest being one of the fitting parameters. Units are kcal mol^{-1} (free energies) or kcal mol^{-1} M^{-1} (*m*-values). All mutants produced in pseudo-WT background of H46N.

that is always structured in the TS but in a non-native binding geometry or a site that is only fractionally populated in the TS with native-like binding affinity. If the site is fully populated but with a non-native binding affinity, the near unity $\phi^{crosslink}$-value observed at this site is expected. For a biHis site to be metal binding-competent, even if in a non-native geometry, it is necessary for the individual histidines to be in proximity. The introduction of a cross-link to this already closed loop would not further reduce the configurational entropy of the TS. As a result, the $\phi^{crosslink}$-value should be unity. Therefore, both measures report that strands $\beta1$ and $\beta2$ are associated in the TS, while the ψ-value conveys additional information about the conformations of the histidine side-chains.

Alternatively, the fractional ψ-value observed for site *n* could be interpreted as a site that is fractionally populated in the TS. The ψ-analysis data at sites *a* and *b*, as well as the current cross-linking results, unequivocally indicate that the hairpin is formed in the TS. Therefore, the loop formed by this hairpin is nearly closed in the TS. If site *n* is fractionally populated in the TS, then the strands are only minimally separated at the open end of the hairpin and only some fraction of the time. Hence, the introduction of a cross-link at the end of the pre-existing hairpin results in only a small loss of entropy in the TS. As a result, the observed $\phi^{crosslink}$ should be near unity for this site, even if this end of the hairpin is only fractionally populated and exhibits a fractional ψ-value. This scenario is diagrammed in Figure 2C.

As just described for the $\beta1$-$\beta2$ hairpin, cross-linking can result in values of $\phi^{crosslink}$ that are greater than their corresponding ψ-values. Generally, cross-linking at a site adjacent to a region that is native-like in the TS represents one such circumstance. Additionally, for a region that is fractionally populated in the TS, the introduction of a cross-link will stabilize the region and increase the fraction of time the biHis site is formed. For example, a cross-link that stabilizes a site by 1.4 kcal mol^{-1} would shift the equilibrium for the formation of that site by a factor of 10. The site would then be more populated in the TS and result in a $\phi^{crosslink}$ that is greater than the corresponding ψ-value.

Metal ion binding experiments performed with src SH3 provide supporting evidence for our interpretation of the underlying mechanism for situations where $\phi^{crosslink} > \psi$-value. As with Ub site *n* on the $\beta1$-$\beta2$ hairpin, the cross-linking results on the RT loop (*18*) indicate that it is more structured than indicated by the metal ion binding results. The ψ-values measured at the base (site *a*) and farther up the loop (site *b*) indicate that some amount of structure is formed toward the turn while the strand ends themselves are not specifically associated. As a result, a near-zero ψ-value is measured for site *a* at the base of the loop, while a fractional $\phi^{crosslink}$-value is measured near this site ($\phi^{crosslink} = 0.45$) due to loop closure providing only an intermediate entropic benefit. This example represents another instance in which cross-linking at an unstructured site adjacent to a partially structured region in the TS results in a $\phi^{crosslink}$-value that is greater than the corresponding ψ-value. In addition, as suggested above and by Baker and co-workers, the stabilization provided by the cross-link shifts the equilibrium, increasing the structure formed in this region in the TSE.

Results for cross-links located at "single loop topology" sites prove the most straightforward to interpret. The $\phi^{crosslink}$-values measured for these sites are identical to or greater than their corresponding ψ-values. Most significantly, the interpretations of the cross-linking results for these sites are consistent with the interpretations for the metal-binding results. Furthermore, the interpretations of the results for "single loop" sites are consistent with the assumption that cross-linking stabilizes the protein by reducing the entropy of the unfolded chain. Cross-linking data at these sites report on the amount of structure contained in the enclosed region, irrespective of whether the loop ends themselves are closed in the TS or not. This is in contrast to metal binding that reports on the interactions of the specific residues at the site examined.

Nonlocal Cross-Links. The remaining cross-linking sites examined are topologically similar to what Thornton terms "nonlocal" cross-links (*26*). These sites cross-link sequence distant residues and close large loops that encompass three or more secondary structural elements. The $\phi^{crosslink}$-values for these sites are mostly smaller than their corresponding ψ-values. Unlike $\phi^{crosslink}$-values that are greater than their corresponding ψ-values, it is more difficult to reconcile these $\phi^{crosslink}$-values, especially within the framework used to interpret the results for the "single loop" sites.

For no site is this difficulty more pronounced than for nonlocal site *g*. The cross-link at this site spans strands $\beta3$ and $\beta4$ and closes a loop that contains over thirty residues and encompasses three β-strands and a small 3_{10} helix. According to the metal ion binding data, this site is fully native-like in the TS with a ψ-value of 1.25 ± 0.04. However, the $\phi^{crosslink}$-value of 0.32 ± 0.01 indicates that far less structure is formed in the TS. Given the unambiguous interpretation of a ψ-value of unity, it is likely that $\phi^{crosslink}$-value for this site is inaccurate or flawed.

One possible explanation for the observed fractional $\phi^{crosslink}$-value is that the cross-link has a destabilizing effect

on the native state, possibly due to conformational strain or steric interference. This effect may be mitigated in the TS by less constrained native-like structures or better accommodated due to TS relaxation. If the native state is destabilized by the introduction of the cross-link, then both the barrier to folding and the barrier to unfolding will be altered. As a result, a fractional $\phi^{crosslink}$-value would be observed, although the region probed is native-like and the loop is closed. This explanation abandons the assumption that the effect of cross-linking is restricted to the unfolded chain and that any effect on native or native-like structures is benign. While the exact nature of the effect is unclear, it seems reasonable to abandon the assumption in this case based on the strength of the interpretation of a ψ-value of unity. Further evidence of the unsuitability of this cross-link for identifying structure in the TS is the altered m_f and m_u values observed in the chevron data (Figure 3 and Table 1).

The $\phi^{crosslink}$-value for a second nonlocal site, site *c*, is also considerably smaller than its corresponding ψ-value (0.12 ± 0.09 versus 0.52 ± 0.03). Bridging strands β2 and β3, site *c* closes the largest loop of any of the sites examined. Despite this fact, site *c* appears to gain little stability as a result of the cross-linking. The small change in stability observed for this site is unusual and suggests that this result is unreliable. The interpretation of the cross-linking result for this site is not easily rationalized and certainly cannot be reconciled with the interpretation of the metal binding result.

The fractional $\phi^{crosslink}$-value for site *h*, the last nonlocal site examined, appears to be consistent with its previously reported ψ-value (0.42 ± 0.04 versus 0.52 ± 0.07). Like site *g*, site *h* cross-links strands β3 and β4. Given the significant changes observed to the chevron m_f- and m_u-values, it seems likely that the apparent consistency of the results of the two methods is merely a coincidence for this site.

Although the sites examined were few, the interpretation of cross-linking data for nonlocal sites poses a greater challenge than the results of the "single loop" sites. The $\phi^{crosslink}$-values measured for two of the three nonlocal sites examined are smaller than their corresponding ψ-values. An explanation for this discrepancy is that these cross-links have an adverse effect on the native state. The exact nature of this adverse effect is not clear. A variety of factors may contribute, including conformational strain or steric interference. Overall, caution is advisable when implementing and interpreting cross-linking results for nonlocal sites.

Comparison with Previous Kinetic Cross-Linking Studies. The results of the current experiments have revealed some potential idiosyncrasies in the way cross-linking reports on the folding TS. It is beneficial to consider the results of previous kinetic studies of cross-linking for corroborating or contradicting evidence. The results of cross-linking across "single loop topology" sites tend to be skewed toward higher $\phi^{crosslink}$-values. For example, a site with a ψ-value of zero in the src SH3 resulted in a fractional $\phi^{crosslink}$-value, while a site with a fractional ψ-value in Ub resulted in a $\phi^{crosslink}$-value near unity. The results of cross-links across the more topologically complicated nonlocal sites, on the other hand, appear to skew toward lower $\phi^{crosslink}$-values. A site with a fractional ψ-value in Ub yielded a $\phi^{crosslink}$-value near zero, while another site with a ψ-value near unity resulted in a fractional $\phi^{crosslink}$-value. If these are general trends, far more $\phi^{crosslink}$-values of unity would be expected at "single loop topology" sites than at the nonlocal sites.

The cross-links in all three of the previously studied proteins, barnase (*10*), src SH3 (*18*), and CD2 (*12*), were achieved by means of engineered disulfide bridges. The linking sites in these proteins were all located across β sheets or flexible loops. Two-thirds (12 of 18 sites) of the sites examined in these proteins match the criteria for a "single loop topology" site. Of the 12 "single loop" sites, six sites had $\phi^{crosslink}$-values of essentially unity while one site had $\phi^{crosslink}$-value of zero. Of the remaining five sites, two (sites *g* and *l* in CD2) were found to destabilize the protein when cross-linked (*12*). A third site in CD2, site *m*, had the equivalent of a fractional $\phi^{crosslink}$-value. Site *m* is adjacent to site *l*, which was one of the two destabilizing cross-links. Although site *m* closes a "single loop", an artifact arising from a destabilized native state could still account for the fractional result at this site. A fractional $\phi^{crosslink}$-value was also observed for the cross-link across the base of the RT loop of src SH3 domain (*11*). However, as already discussed comparison to metal ion binding data for this site reveals that this result fits the pattern observed for sites in Ub and is consistent with our interpretation. The final "single loop" cross-linking site is located between residues 85 and 102 in barnase (*10*). The result for this site is also fractional. The cysteine substitution at residue 102 is located in a β-turn and could potentially affect the native state adversely, resulting in a fractional $\phi^{crosslink}$-value.

In the same three proteins, eight sites were examined that close large, topologically complicated loops. No $\phi^{crosslink}$-values of unity were observed at these sites. Cross-linking at one of the sites, site *b* in CD2, destabilized the protein (*12*). For a "nonlocal" site in barnase, cross-linking resulted in a $\phi^{crosslink}$-values of zero. The cross-links at the remaining sites resulted in fractional $\phi^{crosslink}$-values. Although this survey of past results is not conclusive, the trends observed conform to our expectations. A more extensive comparison of the results of cross-linking to those of metal binding is required to confirm these observations.

CONCLUSION

With the extensive characterization of the long-range contacts formed in Ub's TS using ψ-analysis, we are well positioned to evaluate the utility of cross-linking in identifying structure in the TS. The differences between the two methods primarily arise because of the fundamentally dissimilar means of protein stabilization. Because of its ability to selectively stabilize native-like structure once it is formed, metal ion binding provides a more refined tool for identifying the structure present at a particular site in the TS. Cross-linking is a more coarse-grained tool that appears to be reliable only on local hairpins. In these situations, cross-linking provides information about the loss of configuration entropy of the loop, which is affected by the existence and location of structure enclosed by the cross-link.

ACKNOWLEDGMENT

We thank E. Shakhnovich and members of our group for comments and discussions, Devin Strickland for assistance and advice, and B. Krantz for suggesting the use of DCA as an intramolecular cross-linker.

REFERENCES

1. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, *J. Mol. Biol. 224*, 771−782.
2. Matthews, C. R. (1987) Effect of point mutations on the folding of globular proteins, *Methods Enzymol. 154*, 498−511.
3. Goldenberg, D. P. (1992) Mutational Analysis of Protein Folding and Stability, in *Protein Folding* (Creighton, T. E., Ed.) pp 353−403, W. H. Freeman, New York.
4. Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004) Discerning the structure and energy of multiple transition states in protein folding using psi-analysis, *J. Mol. Biol. 337*, 463−475.
5. Krantz, B. A., and Sosnick, T. R. (2001) Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil, *Nat. Struct. Biol. 8*, 1042−1047.
6. Pandit, A. D., Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2007) Characterizing protein folding transition states using Psi-analysis, *Methods Mol. Biol. 350*, 83−104.
7. Sosnick, T. R., Krantz, B. A., Dothager, R. S., and Baxa, M. (2006) Characterizing the Protein Folding Transition State Using psi Analysis, *Chem. Rev. 106*, 1862−1876.
8. Pandit, A. D., Jha, A., Freed, K. F., and Sosnick, T. R. (2006) Small Proteins Fold Through Transition States With Native-like Topologies, *J. Mol. Biol. 361*, 755−770.
9. Sosnick, T. R., Dothager, R. S., and Krantz, B. A. (2004) Differences in the folding transition state of ubiquitin indicated by phi and psi analyses, *Proc. Natl. Acad. Sci. U.S.A. 101*, 17377−17382.
10. Clarke, J., and Fersht, A. R. (1993) Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation, *Biochemistry 32*, 4322−4329.
11. Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000) Long-range order in the src SH3 folding transition state, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7084−7089.
12. Mason, J. M., Gibbs, N., Sessions, R. B., and Clarke, A. R. (2002) The influence of intramolecular bridges on the dynamics of a protein folding reaction, *Biochemistry 41*, 12093−12099.
13. Jacobson, H., and Stockmayer, W. H. (1950) Intramolecular reactions and polycondensation: I. the theory of linear systems, *J. Chem. Phys. 18*, 1600−1606.
14. Schoenbrunner, N., Pappenberger, G., Scharf, M., Engels, J., and Kiefhaber, T. (1997) Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: Evidence for hairpins as initiation sites for β-sheet formation, *Biochemistry 36*, 9057−9065.
15. Krantz, B. A., and Sosnick, T. R. (2000) Distinguishing between two-state and three-state models for ubiquitin folding, *Biochemistry 39*, 11696−11701.
16. Xu, W., Harrison, S. C., and Eck, M. J. (1997) Three-dimensional structure of the tyrosine kinase c-Src, *Nature 385*, 595−602.
17. Yu, H., Rosen, M. K., and Schreiber, S. L. (1993) 1H and 15N assignments and secondary structure of the Src SH3 domain, *FEBS Lett. 324*, 87−92.
18. Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000) Long-range order in the src SH3 folding transition state, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7084−7089.
19. Yin, L., Krantz, B., Russell, N. S., Deshpande, S., and Wilkinson, K. D. (2000) Nonhydrolyzable diubiquitin analogues are inhibitors of ubiquitin conjugation and deconjugation, *Biochemistry 39*, 10001−10010.
20. Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004) Erratum to Discerning the structure and energy of multiple transition states in protein folding using psi-analysis, *J. Mol. Biol. 347*, 889−1109.
21. Flory, P. J. (1953) *Statistical Mechanics of Chain Molecules*, Cornell University Press, Ithaca, NY.
22. Jacobson, H., Beckmann, C. O., and Stockmayer, W. H. (1950) Intramolecular reaction in polycondensations. II. Ring-chain equilibrium in polydecamethylene adipate, *J. Chem. Phys. 18*, 1607−1612.
23. Kurchan, E., Roder, H., and Bowler, B. E. (2005) Kinetics of loop formation and breakage in the denatured state of iso-1-cytochrome c, *J. Mol. Biol. 353*, 730−743.
24. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding, *Nat. Struct. Biol. 6*, 1016−1024.
25. Northey, J. G., Di, Nardo, A. A., and Davidson, A. R. (2002) Hydrophobic core packing in the SH3 domain folding transition state, *Nat. Struct. Biol. 9*, 126−130.
26. Thornton, J. M. (1981) Disulphide bridges in globular proteins, *J. Mol. Biol. 151*, 261−287.
27. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., Vierstra, R. D., Hatfield, P. M., and Cook, W. J. (1987) Comparison of the three-dimensional structures of human, yeast, and oat ubiquitin, *J. Biol. Chem. 262*, 6396−6399.

BI701042E