

---

**FOR THE RECORD**

# Protein folding transition states probed by loop extension

---

IGNACIO ENRIQUE SÁNCHEZ

Fundación Instituto Leloir, Buenos Aires, Argentina

(RECEIVED September 3, 2007; FINAL REVISION October 17, 2007; ACCEPTED October 17, 2007)

**Abstract**

We propose a new way to characterize protein folding transition states by (1) insertion of one or more residues into an unstructured protein loop, (2) measurement of the effect on protein folding kinetics and thermodynamics, and (3) analysis of the results in terms of a rate-equilibrium free energy relationship,  $\alpha_{\text{Loop}}$ .  $\alpha_{\text{Loop}}$  reports on the fraction of molecules that form the perturbed loop in the transition state. Interpretation of the changes in equilibrium free energy using standard polymer theory can help detect residual structure in the unfolded state. We illustrate our approach with data for the model proteins CI2 and the alpha spectrin SH3 domain.

**Keywords:** protein folding; transition state; loop; rate-equilibrium free energy relationships; kinetics

Protein folding transition states are transient ensembles of conformations that cannot be isolated for their study. Their structure must be characterized by indirect means, commonly rate-equilibrium free energy relationships (REFERs) (Leffler 1952; Sánchez and Kiefhaber 2003c). Briefly, we apply a perturbation  $\partial x$  to a protein folding reaction and measure the changes in the activation free energy for folding,  $\Delta G_f^{0,\ddagger}$ , and the equilibrium free energy,  $\Delta G^0$ . In a two-state folding reaction, these changes are linearly related and a proportionality constant,  $\alpha_x = (\partial \Delta G_f^{0,\ddagger} / \partial x) / (\partial \Delta G^0 / \partial x)$ , can be defined. We can obtain global information about the transition state using medium-induced REFERs and local information using structure-induced REFERs (Sánchez and Kiefhaber 2003c). Local structural perturbations used to date include mutation of residue side chains ( $\phi$ -values) (Fersht et al. 1992), chemical modification of the protein backbone (Deechongkit et al. 2004), and engineering of metal binding

sites ( $\psi$ -values) (Krantz and Sosnick 2001). Commonly, the transition state for folding of a protein is characterized at near-atomic resolution by measuring a large number of  $\phi$ -,  $\psi$ -, or  $\alpha_{\text{Backbone}}$ -values distributed throughout the structure. The  $\phi$ -,  $\psi$ -, and  $\alpha_{\text{Backbone}}$ -values for a given region of the protein are sometimes different (Deechongkit et al. 2004; Sosnick et al. 2004; Pandit et al. 2006), and there is currently no consensus on how to integrate them into a unified description of the transition state (Sosnick et al. 2004; Pandit et al. 2006). This puzzle prompts for the development of additional ways to characterize protein folding transition states.

**Results and Discussion***Characterization of protein folding transition states*

We propose a new structure-induced REFER. Let us consider the simple, two-state folding reaction of a protein that has an unstructured loop. The unstructured loop may be present in the wild-type protein or created by inserting two or more glycine residues into a structured loop. We can then insert one or more extra residues into the already unstructured loop using molecular biology techniques and

---

Reprint requests to: Ignacio Enrique Sánchez, Fundación Instituto Leloir, Avenida Patricias Argentinas, Buenos Aires, Argentina; e-mail: isanchez@leloir.org.ar; fax: 54-11-52387501.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.073217708>.

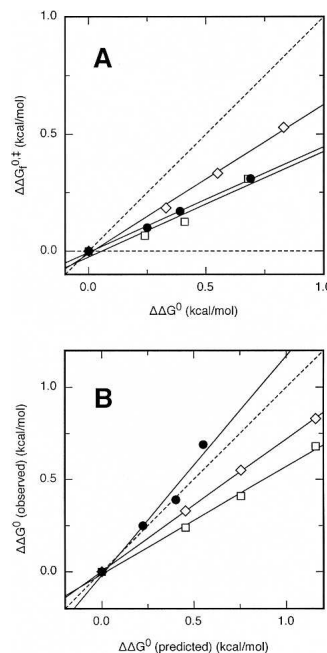
measure the effect on the thermodynamics and kinetics of folding. This perturbation can be understood as a loop extension-induced REFER:

$$\alpha_{\text{Loop}} = \frac{\partial \Delta G_f^{0,\ddagger} / \partial \text{Loop Length}}{\partial \Delta G^0 / \partial \text{Loop Length}}$$

We can give a physical interpretation to  $\alpha_{\text{Loop}}$  using general concepts from polymer theory. The end-to-end distance of the chosen unstructured loop is strongly constrained in the native state and, in the absence of residual structure, unconstrained in the unfolded state. In this case, folding of the loop from the unfolded to the native state should have a free energy cost of purely entropic origin. Extension of the loop will increase this free energy cost because of a larger entropic penalty in the native state and an unchanged (null) entropic penalty in the unfolded state (Jacobson and Stockmeier 1950; Zhou 2004). This will lead to a change in  $\Delta G^0$ . If the loop is constrained into a native-like conformation in the transition state, the change in  $\Delta G_f^{0,\ddagger}$  will be the same as the change in  $\Delta G^0$ , and the  $\alpha_{\text{Loop}}$  will be 1. If the loop is as unconstrained in the transition state as in the unfolded state, the change in  $\Delta G_f^{0,\ddagger}$  will be 0 and the  $\alpha_{\text{Loop}} = 0$ .

Many  $\phi$ -,  $\psi$ -, and  $\alpha_{\text{Backbone}}$ -values measured to date are fractional (Goldenberg 1999; Krantz and Sosnick 2001; Sánchez and Kiefhaber 2003d; Deechongkit et al. 2004; Krantz et al. 2004; Sosnick et al. 2004; Pandit et al. 2006). Unfortunately, their structural interpretation is ambiguous. They may arise from conformational heterogeneity in the transition state, from weakened interactions, or both (Fersht et al. 1992; Bodenreider and Kiefhaber 2005). To a first approximation, the change in free energy caused by extension of an unstructured loop that does not interact with the rest of the protein is independent of other interactions. In this case, the  $\alpha_{\text{Loop}}$  can be read as the fraction of molecules in which the perturbed loop is constrained in the transition state in a native-like manner. However, nonnative-like conformations of the loop in the transition state are also possible. For example, partial opening of the loop neck may change the stiffness of the loop and its response to extension. Alternatively, additional residues may become unstructured and be incorporated into the loop. We can approximate the change in free energy for the addition of  $n$  residues to an unstructured loop of  $N$  residues as  $\Delta \Delta G^0 = 1.5 \cdot RT \cdot \ln(1 + n/N)$  (Jacobson and Stockmeier 1950). It follows that an extended nonnative-like loop in the transition state will have smaller values of  $\Delta G_f^{0,\ddagger}$  and of the  $\alpha_{\text{Loop}}$  than a native-like one. In this case, the  $\alpha_{\text{Loop}}$  gives a lower limit for the fraction of loop-forming molecules (native-like or nonnative-like) in the transition state.

Several protein folding reactions have been perturbed by the extension of an unstructured protein loop. We will use these data to illustrate the use of the  $\alpha_{\text{Loop}}$ . Figure 1A shows the effect of loop extension in the equilibrium and activation free energies for folding of CI2 and the S19P20 and N47D48 circular permutants of the spectrin SH3 domain (Ladurner and Fersht 1997; Viguera and Serrano 1997). The data for the three proteins are well described by a linear fit ( $R = 0.97$  or higher), suggesting that the rate-limiting transition state is not altered by extension of the unstructured loop (Sánchez and Kiefhaber 2003b,a). The slope for each protein, i.e., the  $\alpha_{\text{Loop}}$ -value, varies between 0.45 and 0.64 (Table 1). The fractional  $\alpha_{\text{Loop}}$ -values indicate that the perturbed loops are formed in  $\sim 45\%$ – $64\%$  of the molecules in the transition states for folding of these proteins. The  $\phi$ -values probing side chain interactions across the extended loop are significantly higher than zero in all three proteins (Itzhaki et al. 1995; Viguera et al. 1996), in agreement with our results.



**Figure 1.** (A) Loop extension-induced REFERs for CI2 (●), SH3<sup>S19P20</sup> (□), and SH3<sup>N47D48</sup> (◇) (Ladurner and Fersht 1997; Viguera and Serrano 1997). The continuous lines are linear fits to the data. The dashed lines correspond to  $\alpha_{\text{Loop}}$ -values of 0 and 1 and are shown as a reference. (B) Observed versus predicted effects of loop extension on  $\Delta G^0$  for CI2 (●), SH3<sup>S19P20</sup> (□), and SH3<sup>N47D48</sup> (◇) (Ladurner and Fersht 1997; Viguera and Serrano 1997). The continuous lines are linear fits to the data. The dashed line has a slope of 1 and corresponds to a negligible amount of residual structure in the unfolded state. The predicted  $\Delta \Delta G^0$ -value for the addition of  $n$  residues to an unstructured loop of  $N$  residues is  $\Delta \Delta G^0 = 1.5 \cdot RT \cdot \ln(1 + n/N)$  (Jacobson and Stockmeier 1950).  $N = 7$  and  $n = 2$ – $6$  for CI2;  $N = 3$  and  $n = 2$ – $8$  for both SH3 domains.

**Table 1.** Experimental results for CI2, SH3<sup>S19P20</sup>, and SH3<sup>N47D48</sup>

Protein	N	n	$\alpha_{\text{Loop}}$	Loop structure in U
CI2 <sup>a</sup>	7	2–6	$0.45 \pm 0.01$	$-0.19 \pm 0.15$
SH3 <sup>S19P20a</sup>	3	2–8	$0.45 \pm 0.08$	$0.42 \pm 0.03$
SH3 <sup>N47D48a</sup>	3	2–8	$0.64 \pm 0.03$	$0.28 \pm 0.01$

<sup>a</sup>Ladurner and Fersht (1997) and Viguera and Serrano (1997).

N is the number of residues in the reference unstructured loop and n is the number of inserted residues.  $\alpha_{\text{Loop}}$  is the slope of the plot of  $\Delta\Delta G_{f, \ddagger}^0$  versus  $\Delta\Delta G^0$ . The fraction of loop structure in U is 1 minus the slope of the plot of observed versus predicted  $\Delta\Delta G^0$ -values.

### Detection of residual structure in the unfolded state

All REFERs take the free energy (and, consequently, the structure) of the unfolded state as a reference (Fersht et al. 1992). Many small proteins show substantial residual structure in the unfolded state (Sánchez and Kiefhaber 2003b), which puts a cautionary note to the interpretation of structure-induced REFERs (Cho and Raleigh 2006). Quantification of residual structure at each site probed by a  $\phi$ -,  $\psi$ -, or  $\alpha_{\text{Backbone}}$ -value is time consuming and rarely carried out. The equilibrium data collected in loop extension experiments can be used to detect residual structure in the unfolded state as follows: If a fraction of the protein molecules form the perturbed loop in the unfolded state, they will not pay an entropic penalty upon going from the unfolded to the native state and will not contribute to  $\Delta\Delta G^0$ . The observed  $\Delta\Delta G^0$  will be reduced relative to the one predicted for a fully unfolded state in an amount proportional to the population of the loop. A plot of observed  $\Delta\Delta G^0$ -values versus the ones predicted from polymer theory (Jacobson and Stockmeier 1950) will be linear, with a slope that can be read as the fraction of molecules that do not form a loop in the unfolded state.

Figure 1B shows the observed and predicted  $\Delta\Delta G^0$ -values for CI2, SH3<sup>S19P20</sup>, and SH3<sup>N47D48</sup> (Ladurner and Fersht 1997; Viguera and Serrano 1997). The data for the three proteins are well described by a linear fit ( $R = 0.98$  or higher). The slope for CI2 is  $\sim 1$ , indicating that the two halves of the protein separated by the extended loop, CI2<sup>1–40</sup> and CI2<sup>41–64</sup>, are not in contact in the unfolded state. NMR experiments indeed show that residual structure in the unfolded state is confined to CI2<sup>1–40</sup> (Kazmirski et al. 2001). The slope of such a plot for a series of loop extension variants of the Rop protein (Nagi and Regan 1997) is also very close to 1 (Zhou 2004). On the other hand, the slopes for SH3<sup>S19P20</sup> and SH3<sup>N47D48</sup> are smaller than 1 (Table 1), suggesting the existence of a population with significant residual structure in the unfolded states of both proteins. NMR experiments found extensive residual structure in the unfolded state of the spectrin SH3 domain (Kortemme et al. 2000). This structure involves residues

at both sides of the perturbed loop, in agreement with our results.

### Applicability

In order to yield a meaningful  $\alpha_{\text{Loop}}$ -value, the following conditions must be fulfilled: (1) Creation of an unstructured loop by the insertion of several glycine residues should not change the folding mechanism and (2) should not destabilize the protein drastically, so that the extension variants can be characterized. (3) The inserted sequence should be devoid of interactions with itself or the rest of the protein in the unfolded, native, or transition states. (4) The  $\Delta\Delta G^0$  caused by loop extension should be of entropic origin. The available experimental evidence for creation (Sondek and Shortle 1990; Ladurner and Fersht 1997; Viguera and Serrano 1997; Gordon-Smith et al. 2001) and extension (Martinez et al. 1999) of unstructured loops in the proteins CI2, SH3<sup>S19P20</sup>, SH3<sup>N47D48</sup>, and staphylococcal nuclease suggests that these conditions are often met, and that insertion of polyglycine or polyglutamine sequences into most protein loops will yield a useful  $\alpha_{\text{Loop}}$ -value.

We have shown that loop extension-induced REFERs can give information about loop formation in protein folding transition states.  $\alpha_{\text{Loop}}$  is an approximation for the fraction of loop-forming molecules in the transition state, a piece of information that can be directly compared with simulation. Additionally, interpretation of the results using standard polymer theory can help detect residual structure in the unfolded state. Since only a handful of loops in a protein will be suitable for extension experiments,  $\alpha_{\text{Loop}}$ -values should be used in combination with other medium- and structure-induced REFERs in order to reach a unified description of the transition state.

### Acknowledgments

I am the recipient of an A.E.C.I. fellowship (Spanish Agency for International Cooperation). I would like to acknowledge Gonzalo de Prat-Gay for his mentorship and Diana Wetzler and Mariano Dellarole for comments on the manuscript.

### References

- Bodenreider, C. and Kiefhaber, T. 2005. Interpretation of protein folding  $\psi$  values. *J. Mol. Biol.* **351**: 393–401.
- Cho, J.H. and Raleigh, D.P. 2006. Denatured state effects and the origin of nonclassical  $\phi$  values in protein folding. *J. Am. Chem. Soc.* **128**: 16492–16493.
- Deechongkit, S., Nguyen, H., Powers, E.T., Dawson, P.E., Gruebele, M., and Kelly, J.W. 2004. Context-dependent contributions of backbone hydrogen bonding to  $\beta$ -sheet folding energetics. *Nature* **430**: 101–105.
- 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.
- Goldenberg, D.P. 1999. Finding the right fold. *Nat. Struct. Biol.* **6**: 987–990.

- Gordon-Smith, D.J., Carbajo, R.J., Stott, K., and Neuhaus, D. 2001. Solution studies of chymotrypsin inhibitor-2 glutamine insertion mutants show no interglutamine interactions. *Biochem. Biophys. Res. Commun.* **280**: 855–860.
- Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* **254**: 260–288.
- Jacobson, H. and Stockmeier, W.H. 1950. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* **18**: 1600–1606.
- Kazmirski, S.L., Wong, K.B., Freund, S.M., Tan, Y.J., Fersht, A.R., and Daggett, V. 2001. Protein folding from a highly disordered denatured state: The folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl. Acad. Sci.* **98**: 4349–4354.
- Kortemme, T., Kelly, M.J., Kay, L.E., Forman-Kay, J., and Serrano, L. 2000. Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **297**: 1217–1229.
- 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.
- 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.
- Ladurner, A.G. and Fersht, A.R. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J. Mol. Biol.* **273**: 330–337.
- Leffler, J.E. 1952. Parameters for the description of transition states. *Science* **117**: 340–341.
- Martinez, J.C., Viguera, A.R., Berisio, R., Wilmanns, M., Mateo, P.L., Filimonov, V.V., and Serrano, L. 1999. Thermodynamic analysis of  $\alpha$ -spectrin SH3 and two of its circular permutants with different loop lengths: Discerning the reasons for rapid folding in proteins. *Biochemistry* **38**: 549–559.
- Nagi, A.D. and Regan, L. 1997. An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold. Des.* **2**: 67–75.
- 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.
- Sánchez, I.E. and Kiefhaber, T. 2003a. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**: 367–376.
- Sánchez, I.E. and Kiefhaber, T. 2003b. Hammond behavior versus ground state effects in protein folding: Evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **327**: 867–884.
- Sánchez, I.E. and Kiefhaber, T. 2003c. Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys. Chem.* **100**: 397–407.
- Sánchez, I.E. and Kiefhaber, T. 2003d. Origin of unusual  $\phi$ -values in protein folding: Evidence against specific nucleation sites. *J. Mol. Biol.* **334**: 1077–1085.
- Sondek, J. and Shortle, D. 1990. Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. *Proteins* **7**: 299–305.
- 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.* **101**: 17377–17382.
- Viguera, A.R. and Serrano, L. 1997. Loop length, intramolecular diffusion and protein folding. *Nat. Struct. Biol.* **4**: 939–946.
- Viguera, A.R., Serrano, L., and Wilmanns, M. 1996. Different folding transition states may result in the same native structure. *Nat. Struct. Biol.* **3**: 874–880.
- Zhou, H.X. 2004. Loops, linkages, rings, catenanes, cages, and crowders: Entropy-based strategies for stabilizing proteins. *Acc. Chem. Res.* **37**: 123–130.