

Definition of amide protection factors for early kinetic intermediates in protein folding

WALID A. HOURY*[†], J. MICHAEL SAUDER^{‡§}, HEINRICH RODER^{‡§}, AND HAROLD A. SCHERAGA*[¶]

*Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853-1301; [†]Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111; and [§]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059

Contributed by Harold A. Scheraga, January 29, 1998

ABSTRACT Hydrogen–deuterium exchange experiments have been used previously to investigate the structures of well defined states of a given protein. These include the native state, the unfolded state, and any intermediates that can be stably populated at equilibrium. More recently, the hydrogen–deuterium exchange technique has been applied in kinetic labeling experiments to probe the structures of transiently formed intermediates on the kinetic folding pathway of a given protein. From these equilibrium and nonequilibrium studies, protection factors are usually obtained. These protection factors are defined as the ratio of the rate of exchange of a given backbone amide when it is in a fully solvent-exposed state (usually obtained from model peptides) to the rate of exchange of that amide in some state of the protein or in some intermediate on the folding pathway of the protein. This definition is straightforward for the case of equilibrium studies; however, it is less clear-cut for the case of transient kinetic intermediates. To clarify the concept for the case of burst-phase intermediates, we have introduced and mathematically defined two different types of protection factors: one is P_{struc} , which is more related to the structure of the intermediate, and the other is P_{app} , which is more related to the stability of the intermediate. Kinetic hydrogen–deuterium exchange data from disulfide-intact ribonuclease A and from cytochrome *c* are discussed to explain the use and implications of these two definitions.

Protein folding pathways traditionally have been described as a series of kinetic intermediates that occur before the formation of the native state (1–4). Intermediates that are of most interest are those that form early in the pathway, because they provide insight into the forces that restrict the conformational space of the unfolded protein at the earliest points in the folding process. These conformational preferences are thought to direct folding along a preferred pathway and thus may contribute to the final native structure attained by that protein. However, the structural properties and kinetic roles of these marginally stable structures are a matter of controversy (3–10).

Evidence for early folding events often is based on spectral changes occurring during the dead time of a fast-mixing experiment (typically a few milliseconds), which can be attributed to the rapid formation of an ensemble of compact states. Although optical spectroscopy can provide useful information on the stability and overall conformational properties of these so-called burst-phase intermediates (11–14), their transient nature and dynamic properties have made it difficult to obtain more detailed structural information. The most promising approach makes use of hydrogen–deuterium exchange labeling on a quenched-flow apparatus with subsequent NMR analysis to probe the formation of hydrogen-bonded structure

during folding (15–17). Such hydrogen-exchange competition and pulse-labeling experiments have been the main source of structural information on folding intermediates (reviewed in refs. 18–21).

In a typical pulse labeling experiment, the protein is unfolded in D₂O and allowed to refold under conditions where hydrogen exchange is slow (relatively low pH). After a given refolding time, the protein is briefly exposed to H₂O buffer at basic pH, where accessible amide groups are rapidly protonated while protected ones remain deuterated. The exchange reaction is then quenched by lowering the pH, and the protein is allowed to refold to the native state. The degree of protonation for individual amide sites is determined by two-dimensional NMR analysis of the refolded protein. By performing a series of such labeling experiments at variable refolding times, it is possible to follow the time course of hydrogen-bond formation during folding for many individual amide groups (all those that are well protected from exchange under native conditions). The pulse-labeling protocol also can be extended to measure protection factors in folding intermediates, which provides unique insight into (local) structural stability (22, 23). Exchange rates in transient intermediates can be measured by systematic variation of the labeling conditions (e.g., a short exchange period at variable pH) at a constant refolding time, and protection factors can be estimated by comparison with the corresponding free-peptide exchange rates (24). In a modified version of this protocol, the labeling pulse is applied directly without a refolding delay, which makes it possible to characterize marginally stable structures that accumulate during the dead time of a quenched flow experiment (25). As in the original competition experiment (15, 26), this burst-phase (or dead-time) labeling method makes use of the competition between refolding and H/D exchange to probe structure formation during early stages of folding.

Protein hydrogen exchange results usually are expressed in terms of protection factors, $P = k_{ex}(U)/k_{ex}(N)$, where $k_{ex}(N)$ and $k_{ex}(U)$ represent the rates of exchange for a given amide proton under folded and unfolded conditions, respectively. The exchange rates of fully solvent-exposed amide protons, $k_{ex}(U)$, and their dependence on primary structure and solvent conditions are well understood on the basis of model peptide studies (24, 27) and direct measurements on unfolded proteins (28, 29). Although the definition and interpretation of protection factors for conventional exchange studies on folded proteins under equilibrium conditions are straightforward, this concept is less clear-cut in the case of transient states observed in nonequilibrium labeling experiments. In this communication, we discuss the use and definition of protection factors as applied to burst-phase intermediates.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/954299-4\$2.00/0
PNAS is available online at <http://www.pnas.org>.

[†]Present address: Max-Planck-Institute for Biochemistry, Department of Cellular Biochemistry, Am Klopferspitz 18A, D-82152 Martinsried, Germany.

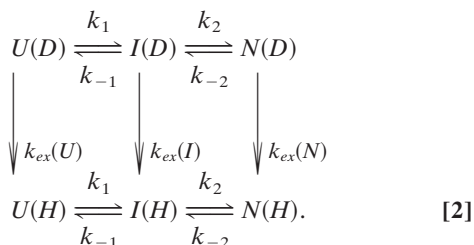
[¶]To whom reprint requests should be addressed. e-mail: has5@cornell.edu.

Definition of Protection Factor

For a general discussion of protection factors, we will assume that the refolding pathway of a hypothetical protein can be described by a simple sequential mechanism in which one burst-phase intermediate is populated. Hence, we can write



U and N are the unfolded and the native states of the protein, respectively, and I represents an ensemble of intermediate states sequentially populated on the refolding pathway. Without making any assumptions concerning the relative rates of exchange and folding, the most general mechanism for the exchange during the labeling pulse of a hydrogen-exchange labeling experiment can be described as follows,



The upper row of the kinetic model represents the refolding of the deuterated protein [indicated by (D)], and the lower row represents the refolding of the protonated protein [indicated by (H)]. The rate constants k_x and k_{-x} represent the forward and backward kinetic rates, and the rate constants $k_{ex}(X)$ represent the rates for *direct* exchange from the deuterated species $X(D)$ to the protonated species $X(H)$, probably through local unfolding events.

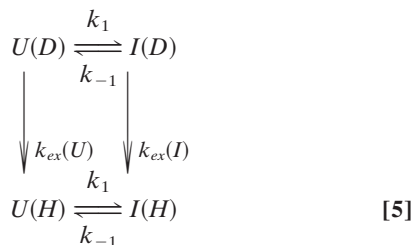
In describing the protection factors in the burst-phase intermediate I , we define the following two quantities: a structural protection factor,

$$P_{struc}(I) = k_{ex}(U)/k_{ex}(I) \quad [3]$$

and an apparent protection factor,

$$P_{app}(I) = k_{ex}(U)/k_{app}(I), \quad [4]$$

where $k_{app}(I)$ represents the apparent rate for the appearance of label in $I(H)$ from $I(D)$ through *all* available pathways including direct exchange and unfolding processes and *excluding* subsequent folding reactions from the burst-phase intermediate. In other words, exchange can take place from the intermediate or from any other state that is more unfolded than the intermediate. Hence, $k_{app}(I)$ is derived from the following part of the mechanism of Eq. 2,



using the following equation,

$$d/dt [I(H)] = [I(D)] k_{app}(I). \quad [6]$$

The following general relation can be helpful in calculating k_{app} when dealing with burst-phase intermediates that are in rapid preequilibrium,

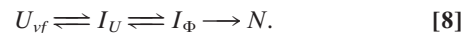
$$d/dt\{[U(H)] + [I(H)] + [N(H)]\} =$$

$$[U(D)] k_{ex}(U) + [I(D)] k_{ex}(I) + [N(D)] k_{ex}(N). \quad [7]$$

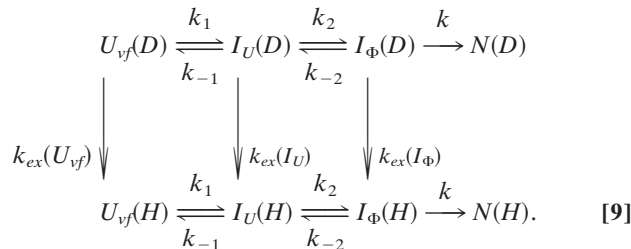
P_{struc} describes the structure in the intermediate irrespective of the state from which the intermediate folded or toward which the intermediate is proceeding. In other words, P_{struc} provides information about the structural stability of a hypothetical equilibrium analogue of the intermediate that is removed from its folding/unfolding pathway. On the other hand, P_{app} more appropriately describes the structural stability of the intermediate in the context of its folding/unfolding pathway. For the case of a protein in an equilibrium state, a protection factor is simply defined as the ratio of the intrinsic (statistical coil or unfolded) exchange rate to the net exchange rate through *all* available pathways, including global and local *unfolding* events. Hence, the apparent protection factor for kinetic intermediates is most similar to the protection factor as defined for equilibrium states. We point out that P_{struc} will always be larger than P_{app} .

Specific Example 1: Ribonuclease A

To provide further understanding of the two definitions of protection factors for kinetic intermediates, we will use the data of Houry and Scheraga (30) as an example. The refolding pathway of the very-fast phase of disulfide-intact ribonuclease A (RNase A) can best be described by the following model,



U_{vf} is the very-fast refolding species, whereas I_U is a largely unfolded intermediate and I_Φ is a hydrophobically collapsed intermediate. Both I_U and I_Φ are burst-phase intermediates that are formed within the dead-time of the stopped-flow instrument. We have carried out pulse-labeling experiments on the refolding pathway of U_{vf} to determine the structure present in I_U and in I_Φ . The experiment is mainly a competition between folding and exchange. The kinetic model that was used to fit the exchange data is as follows,



No direct exchange from N is assumed to occur within the time scale of the labeling pulse (which was 20 ms). Furthermore, any back reactions from N are assumed to be negligible under the conditions of the experiment.

By fitting the above model (Eq. 9), we calculated the exchange rates with respect to that of the unfolded species (U_{vf}), and we defined the protection factor as being the ratio of the exchange rate from U_{vf} [$k_{ex}(U_{vf})$] to that from the intermediate [$k_{ex}(I_U)$ or $k_{ex}(I_\Phi)$]. Such a definition would give the structural protection factor (P_{struc} , Eq. 3) and *not* the apparent protection factor (P_{app} , Eq. 4).

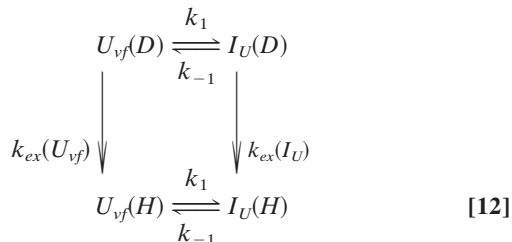
We now derive the relation between P_{struc} and P_{app} for the case of refolding of U_{vf} (Eq. 9). As discussed by Houry and Scheraga (30), the rapid equilibrium between U_{vf} , I_U , and I_Φ is established before formation of N . In other words, k_1 , k_{-1} , k_2 , and $k_{-2} \gg k$. Hence, we can write at any time during the labeling pulse,

$$K_i = k_1/k_{-1} = [I_U(H)]/[U_{vf}(H)] = [I_U(D)]/[U_{vf}(D)] \quad [10]$$

and

$$K = k_2/k_{-2} = [I_\Phi(H)]/[I_U(H)] = [I_\Phi(D)]/[I_U(D)]. \quad [11]$$

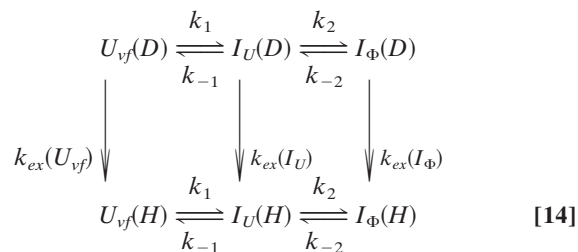
In addition, we have argued that $K_i \gg 1$ and that $K_i \gg K$. We also calculated that $P_{struc}(I_U) \sim 1$. $k_{app}(I_U)$ is derived from the following part of Eq. 9,



using the equation,

$$d/dt [I_U(H)] = [I_U(D)] k_{app}(I_U), \quad [13]$$

and $k_{app}(I_\Phi)$ is derived from the following part of Eq. 9,



using the equation,

$$d/dt [I_\Phi(H)] = [I_\Phi(D)] k_{app}(I_\Phi). \quad [15]$$

Hence, by using Eqs. 12–15 and the relevant form of Eq. 7, together with the above approximations, it can easily be shown that $P_{app}(I_U) = P_{struc}(I_U) = 1$, and

$$\begin{aligned} P_{app}(I_\Phi) &= [P_{struc}(I_\Phi) \cdot P_{struc}(I_U) \cdot (1 + K)]/[P_{struc}(I_\Phi) \\ &\quad + P_{struc}(I_U) \cdot K] \\ &= [P_{struc}(I_\Phi) \cdot (1 + K)]/[P_{struc}(I_\Phi) + K]. \end{aligned} \quad [16]$$

If we consider the limiting case in which there is no direct exchange from I_Φ , i.e., $P_{struc}(I_\Phi) \rightarrow \infty$, then

$$P_{app}(I_\Phi) = (1 + K). \quad [17]$$

If we consider the other limiting case in which there is no protection in I_Φ , i.e., $P_{struc}(I_\Phi) = 1$, then $P_{app}(I_\Phi) = 1$.

In Houry and Scheraga (30), four categories of protected amides were found in I_Φ . As mentioned earlier, the protection factors used were defined as P_{struc} and not as P_{app} . Under the conditions of the experiment, $K = 1.3$. Hence, by using Eq. 16 to relate P_{struc} and P_{app} , the four categories are:

$$P_{struc}(I_\Phi) < 1.5 \Rightarrow P_{app}(I_\Phi) < 1.2 \quad [18a]$$

$$1.5 < P_{struc}(I_\Phi) < 5 \Rightarrow 1.2 < P_{app}(I_\Phi) < 1.8 \quad [18b]$$

$$5 < P_{struc}(I_\Phi) < 50 \Rightarrow 1.8 < P_{app}(I_\Phi) < 2.2 \quad [18c]$$

$$P_{struc}(I_\Phi) > 100 \Rightarrow P_{app}(I_\Phi) \sim 2.3. \quad [18d]$$

It is important to note that, according to Eq. 17, the maximum value for $P_{app}(I_\Phi)$ is 2.3. The range of values for P_{struc} is much wider than that for P_{app} . Consequently, from Eqs. 18a–18d, it can be observed that the information about the structure in I_Φ is less clear when presented in terms of P_{app} .

As can be seen from the above treatment, the mathematical interpretation of P_{app} is much more involved than that of P_{struc} . When the values of $k_{ex}(I)$ can be obtained directly, as in the case of the study by Houry and Scheraga (30), then it is easier to use P_{struc} to gain insight into the structure present in the intermediates. However, in experiments in which the proton occupancies for the amide protons for a well defined intermediate are measured at different refolding times, the exchange rates obtained are those of k_{app} and, consequently, P_{app} is usually calculated to gain insight into the stability of the structure in that intermediate.

Specific Example 2: Cytochrome *c*

In a recent hydrogen-exchange labeling study on horse cytochrome *c* (cyt *c*), J. M. Sauder and H.R. (unpublished data) measured protection factors for about 40 individual amide groups in an early folding intermediate by using the burst-phase competition protocol (25, 31). The competition between hydrogen–deuterium exchange and structure formation during the 2-ms dead time of the quenched-flow experiment was measured as a function of labeling pH under two sets of conditions: (i) strongly native conditions (0.3 M GuHCl, 0.4 M Na₂SO₄) favoring accumulation of early intermediates (32, 33), and (ii) marginally native conditions (2.5 M GuHCl, 0.4 M Na₂SO₄) in which intermediates are destabilized and the native state is formed in an approximate two-state transition. The latter was confirmed by fluorescence-detected stopped-flow measurements of the kinetics of folding, which further showed that the stability of the burst-phase intermediate was essentially independent of pH over the range of 7–11 despite significant destabilization of the native state at alkaline pH.

The burst-phase labeling results were interpreted on the basis of a three-state folding mechanism (Eq. 2), where formation of the native structure (or more structured late intermediates) is preceded by a rapid (<ms) preequilibrium involving a marginally stable intermediate, *I*. Apparent protection factors in *I*, $P_{app}(I)$, were calculated according to Eq. 4, by using the parameters of Bai *et al.* (24) to estimate $k_{ex}(U)$. Under destabilizing conditions (2.5 M GuHCl), the pH profiles for 37 of the 40 amide protons observed were in close agreement with the behavior predicted for a statistical coil, yielding $P_{app}(I) = 1$. However, the pH profiles for three amide protons (Cys-14, Ala-15, and His-18) were shifted toward basic pH by 1–2.5 pH units, indicating apparent protection factors in the range of 10–300. This observation is consistent with earlier evidence for the presence of residual structure in the vicinity of the His-18 iron ligand and the heme attachment site (Cys-14 and Cys-17), which persists even in the fully denatured state (34). The same three amides also showed unusually high protection under stable conditions (0.3 M GuHCl). To correct for this residual structure effect, the apparent protection factors at 0.3 M GuHCl were divided by those at 2.5 M GuHCl. These corrected protection factors are a measure of the net stabilization from formation of intermediates within the first 2 ms of refolding, independent of residual structure in the unfolded state.

Amide groups with significant protection, $P_{app}(I) \geq 2$, were found primarily in three segments of the backbone (residues 9–14, 60–70, and 94–101), which span the major α -helices of the native structure (35). Several helical core residues (F10, M65, L96, K99, A101) showed protection factors in the range of 5–7, which corresponds closely to the maximum protection, $P_{app}(I) = K + 1$ [assuming $k_{ex}(I) = 0$; see Eq. 17], predicted on the basis of the $I \leftrightarrow U$ equilibrium constant, $K \sim 4$, obtained from the fluorescence-detected burst-phase amplitude (J. M. Sauder and H.R., unpublished data). These observations are consistent with the rapid (<2 ms) formation of a marginally stable intermediate with persistent hydrogen-bonded structure limited to three mutually interacting α -helices. A number of

amide groups (especially at the ends of α -helical segments) showed intermediate protection factors in the range of 2–5, indicating that exchange at these sites can occur without complete unfolding of the intermediate. According to Eq. 2, these correspond to sites where $k_{ex}(I) > 0$ [finite $P_{struc}(I)$]. Alternatively, such intermediate protection factors might be interpreted in terms of an expanded mechanism containing additional, less structured intermediates. However, a unique mechanism cannot be obtained, because it is unclear whether such intermediates should be placed on a direct path toward U or off-pathway (36).

The comparison of apparent protection factors with pre-equilibrium measurements relies on the assumption that $k_1 \gg k_{ex}(U)$ (EX2 exchange; ref. 37), which is consistent with the sigmoidal pH dependence for most amide protons observed in the study of J. M. Sauder and H.R. (unpublished data). Moreover, M. C. R. Shastry and H.R. (unpublished data) recently were able to measure the initial rate of folding in cyt *c* directly, by using a continuous-flow mixing method with submillisecond time resolution (38). The rate observed ($2 \times 10^4 \text{ s}^{-1}$) indicates that the EX2 condition should be satisfied up to pH 11.

Conclusions

The definition of protection factors for transient kinetic intermediates is much more involved than that for well defined equilibrium states. We have elucidated mathematically the definition and use of two different types of protection factors for kinetic intermediates, P_{struc} and P_{app} . P_{struc} relates more directly to the structure of the intermediates independent of the folding/unfolding processes, and P_{app} is more closely related to the stability of the intermediates. The relation between P_{struc} and P_{app} can easily be derived for the case of burst-phase intermediates in rapid preequilibrium with the unfolded species.

The definition of P_{struc} implies the presence of direct-exchange mechanisms from the intermediate of interest. The exchange mechanisms are not well understood at this time and most probably involve local unfolding events. By varying the time, pH, and solvent conditions of the pulse, more insight might be gained about these exchange mechanisms. Such information may provide insight into the question of whether protein folding proceeds along well defined pathways populating well defined intermediates, or whether protein folding is better described by funnel-shaped energy landscapes (39, 40). In the latter case, P_{struc} would reflect the exchange from an ensemble of structures present along a folding funnel. Because it would be less clear-cut to define a stability for those structures, the P_{struc} would be a more useful parameter to use than P_{app} .

This work was supported by National Institutes of Health Grants GM-24893 (to H.A.S.), and GM-35926 and CA-06927 (to H.R.), and by an appropriation from the Commonwealth of Pennsylvania to the Institute for Cancer Research.

1. Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660.

2. Matthews, C. R. (1993) *Annu. Rev. Biochem.* **62**, 653–683.
3. Roder, H. & Colón, W. (1997) *Curr. Opin. Struct. Biol.* **7**, 15–28.
4. Rothwarf, D. M., Li, Y.-J. & Scheraga, H. A. (1998) *Biochemistry*, in press.
5. Scheraga, H. A., Konishi, Y., Rothwarf, D. M. & Mui, P. W. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5740–5744.
6. Baldwin, R. L. (1991) *Chemtracts Biochem. Mol. Biol.* **2**, 379–389.
7. Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10869–10873.
8. Creighton, T. E. (1997) *Trends Biochem. Sci.* **22**, 6–11.
9. Eaton, W. A., Muñoz, V., Thompson, P. A., Chan, C.-K. & Hofrichter, J. (1997) *Curr. Opin. Struct. Biol.* **7**, 10–14.
10. Sosnick, T. R., Shitlerman, M. D., Mayne, L. & Englander, S. W. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 8545–8550.
11. Houry, W. A., Rothwarf, D. M. & Scheraga, H. A. (1994) *Biochemistry* **33**, 2516–2530.
12. Roder, H. & Elöve, G. A. (1994) in *Mechanisms of Protein Folding: Frontiers in Molecular Biology*, ed. Pain, R. H. (Oxford Univ. Press, New York), pp. 26–54.
13. Evans, P. A. & Radford, S. E. (1994) *Curr. Opin. Struct. Biol.* **4**, 100–106.
14. Kuwajima, K. (1996) in *Circular Dichroism and the Conformational Analysis of Biomolecules*, ed. Fasman, G. D. (Plenum, New York), pp. 159–182.
15. Roder, H. & Wüthrich, K. (1986) *Proteins* **1**, 34–42.
16. Udgaonkar, J. B. & Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699.
17. Roder, H., Elöve, G. A. & Englander, S. W. (1988) *Nature (London)* **335**, 700–704.
18. Roder, H. (1989) *Methods Enzymol.* **176**, 446–473.
19. Baldwin, R. L. & Roder, H. (1991) *Curr. Biol.* **1**, 218–220.
20. Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 84–91.
21. Englander, S. W. & Mayne, L. (1992) *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243–265.
22. Elöve, G. A. & Roder, H. (1991) *ACS Symp. Ser.* **470**, 50–63.
23. Udgaonkar, J. B. & Baldwin, R. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8197–8201.
24. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993) *Proteins* **17**, 75–86.
25. Gladwin, S. T. & Evans, P. A. (1996) *Folding Des.* **1**, 407–417.
26. Schmid, F. X. & Baldwin, R. L. (1979) *J. Mol. Biol.* **135**, 199–215.
27. Connelly, G. P., Bai, Y., Jen, M.-F. & Englander, S. W. (1993) *Proteins* **17**, 87–92.
28. Roder, H., Wagner, G. & Wüthrich, K. (1985) *Biochemistry* **24**, 7407–7411.
29. Robertson, A. D. & Baldwin, R. L. (1991) *Biochemistry* **30**, 9907–9914.
30. Houry, W. A. & Scheraga, H. A. (1996) *Biochemistry* **35**, 11734–11746.
31. Roder, H., Wand, A. J., Milne, J. S. & Englander, S. W. (1986) *Biophys. J.* **49**, 57a.
32. Elöve, G. A., Chaffotte, A. F., Roder, H. & Goldberg, M. E. (1992) *Biochemistry* **31**, 6876–6883.
33. Colón, W., Elöve, G. A., Wakem, L. P., Sherman, F. & Roder, H. (1996) *Biochemistry* **35**, 5538–5549.
34. Elöve, G. A., Bhuyan, A. K. & Roder, H. (1994) *Biochemistry* **33**, 6925–6935.
35. Bushnell, G. W., Louie, G. V. & Brayer, G. D. (1990) *J. Mol. Biol.* **214**, 585–595.
36. Clarke, J., Itzhaki, L. S. & Fersht, A. R. (1997) *Trends Biochem. Sci.* **22**, 284–287.
37. Hvidt, A. & Nielsen, S. O. (1966) *Adv. Protein Chem.* **21**, 287–386.
38. Shastry, M. C. R., Luck, S. D. & Roder, H. (1998) *Biophys. J.*, in press.
39. Baldwin, R. L. (1995) *J. Biomol. NMR* **5**, 103–109.
40. Dill, K. A. & Chan, H. S. (1997) *Nat. Struct. Biol.* **4**, 10–19.