

Characterizing protein folding intermediates

Exchange rates of individual amide protons in the polypeptide backbone of a protein can be measured by two-dimensional NMR. They serve to locate structure in folding intermediates.

To determine the pathway by which a protein folds up, it is necessary to characterize the structures of folding intermediates and also to place these intermediates in the correct order on the kinetic pathway of folding. Noncovalent folding reactions are fast: typically they occur in seconds or less for small single-domain proteins. On the other hand, hours are required to obtain detailed structural information about a protein in solution by two-dimensional NMR spectroscopy. For proteins whose structures are stabilized by disulfide bonds formed during the folding process, this dilemma can be sidestepped by covalently blocking unreacted thiol groups after partial oxidation of the cysteine residues to form disulfides, and then isolating and characterizing individual disulfide intermediates [1]. In doing this, it is necessary to relate the covalent chemistry of disulfide formation, as successive disulfide bonds are formed, to the noncovalent folding process that is stabilized by H-bonds, hydrophobic interactions, and other weak interactions.

The pulsed NH exchange, or pulse labeling, method [2–6] resolves the dilemma by using fast isotopic exchange ($^2\text{H}\rightarrow^1\text{H}$) between solvent protons and peptide NH (amide) protons to pulse label the accessible amide protons of a noncovalent folding intermediate. Exchange of a solvent-exposed amide proton takes only 10^{-3} seconds at pH 9 and 10°C . Isotopic exchange need not interfere with the folding process, and the results of an exchange pulse are measured after folding is complete, by recording a two-dimensional ^1H -NMR spectrum of the protein, in which ^2H is silent. In its simplest form, the pulse labeling method can be used to locate the protected amide protons, which fail to exchange when the amide protons that are exposed to solvent do exchange.

As Creighton has recently pointed out in this journal [7], this application of the pulse labeling method encounters the following problem. Only a fraction of the peptide NH protons of a protein can be studied because the ^1H -NMR spectrum which is used to determine the results of pulse exchange must be taken in $^2\text{H}_2\text{O}$. Thus, only amide protons that are resistant to exchange with $^2\text{H}_2\text{O}$ in the native protein during the NMR experiment (usually a third to a half of all the amide protons) can be monitored in the process of folding. Often these stable amide protons are found in α -helices and β -strands of the native protein, because the peptide H-bonds of these cooperatively formed structures must be broken to effect exchange. If the available probes happen to be located only in the α -helices and β -structures of the native protein, and if they are all protected in a folding intermediate, then knowledge of their locations in the amino-acid sequence

does not distinguish between different mechanisms of protection: the formation of native-like secondary structure, the formation of a non-native structure that breaks down again later in folding, or protection by a nonspecific mechanism such as a hydrophobic collapse.

The problem is not insoluble. A simple solution is not just to observe the locations of protected amide protons but also to measure their exchange rates, or protection factors, in the folding intermediate. Then the plot of protection factor versus residue position for the folding intermediate can be compared with that of the native protein, to find out if there is evidence for native-like secondary structure in the intermediate. This has been done recently for two equilibrium folding intermediates, as discussed below. The protection factor P_i of proton i gives the retardation of its exchange rate in the folding intermediate, expressed as the ratio of two rates: $P_i = (k_c/k_{ex})_i$, where k_c is the 'chemical' exchange rate of the solvent-exposed proton i , computed from model compound data, and k_{ex} is the experimentally observed rate of exchange in the folding intermediate.

In native proteins, α -helices and β -sheets give rise to characteristic patterns of exchange rates. As the locations of α -helices and β -strands in the amino-acid sequence are also highly specific for each protein, the plot of protection factor versus residue position serves as a fingerprint that can be used to determine if the structure of a folding intermediate is closely related to that of the native protein. Not all amide protons form peptide H-bonds in α -helices and β -strands. The NH protons of the four residues at the amino-terminal end of a regular α -helix do not participate in intra-helix H-bonds and others are involved in irregular tertiary interactions. The β -strands of a protein form a highly specific pattern of H-bonds: for example, the amide protons in the outside strands alternate between being and not being H-bonded within the β -strand structure.

The distinctive relation between H-bonded secondary structure and protection against exchange measured at 36°C is shown in Fig. 1 for bovine pancreatic trypsin inhibitor (BPTI) [8]. Because the exchange rates of BPTI were measured without using a quench, it was possible to measure them for all amide protons that exchange at measurable rates. BPTI is a small 58-residue protein that contains an α -helix with peptide H-bonds from residues 47–57 and a hairpin loop of antiparallel β -structure from residues 16–36 with a short third strand near residue 45. There is a block of six protected amide protons in the α -helix, another block of five very strongly protected protons in the center strand of the β -sheet, and a striking pat-

tern of alternating strong and weak protection for amide protons in the exposed part of the β -hairpin loop.

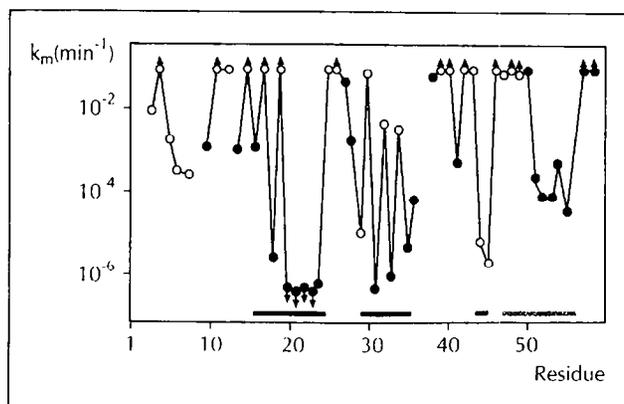


Fig. 1. Exchange rates of backbone amide protons in BPTI plotted against residue number. Amide protons that H-bond are coloured black and the locations of β -strands (green) and an α -helix (orange) are shown. Exchange rates that are too large or too small to measure are shown by red arrows pointing up or down, respectively. (Adapted from [8].)

Now consider the plot of protection factor versus residue position for two equilibrium folding intermediates, cytochrome *c* at pH 2.2 [9] and apomyoglobin at pH 4.2 [10]. The term equilibrium folding intermediate is used here in the sense of a partly folded structure, intermediate between the structures of the native and unfolded protein, and does not imply that the equilibrium intermediate is populated in the kinetic process of folding. Both cytochrome *c* and myoglobin contain α -helices but no β -strand structures. In cytochrome *c*, 27 of the 44 available protons are found within three major α -helices. In native (pH 7) cytochrome *c* (N), the protons within the three helices are strongly protected ($P > 1000$) at 20°C, as are several protons in loop regions (Fig. 2a). In the equilibrium folding intermediate of cytochrome *c* all protons are much less protected than in N, but now there are boundaries between clusters of protons that show marked protection and others that do not; taking $P > 10$ as the criterion for marked protection, one finds a cluster of protected protons corresponding to each of the three α -helices of N. The results suggest that helical structure provides the marked protection and that the boundaries of helices in I are nearly the same as those in N.

A different situation is found in apomyoglobin at 5°C. In N (pH 6.0), only two of the 39 available protons occur outside helices. This is the situation that was feared [7]: if all these protons are protected in I, then a meaningless correlation might be drawn between the locations of helices in N and clusters of protected protons in I. The results, however, show that not all probes in I are protected. Instead, clusters of probes are found, some of which show marked protection whereas others do not (Fig. 2b). Again, using $P > 10$ as the dividing line, one finds clusters of protons that show only borderline protection at the locations of the B and E helices of N. The results for the E helix are particularly striking: the largest value of P is 2. On the other hand, for most protons in the A, G and H helices $P > 10$. Two probes within the G he-

lix with $P < 10$ are among the four residues at the amino terminus that are not H-bonded if the α -helix is regular. Consequently, in apomyoglobin, as in cytochrome *c*, the pattern of protection factors in I provides a fingerprint resembling that of N in the sense that zones of marked protection in I correspond to the locations of helices in N. Although the boundaries of the protected zones in I correspond to the helices of N, there are clear differences in the protection patterns within helices between N and I — see, for example, the G and H helices of myoglobin (Fig. 2b). These differences probably arise from different tertiary interactions in N and I.

The question then arises: do isolated peptides whose sequences correspond to these helices also show strong enough helix formation to provide marked protection against exchange? This question has very recently been answered with synthetic peptides corresponding to the G and H helices of myoglobin [10] and the N and C helices of cytochrome *c* (S Marqusee, D Barrick and R Baldwin, unpublished data; see also [11]). In each case the answer is no. Thus, the structure of the folding intermediate stabilizes the three helices of cytochrome *c* and the A, G and H helices of myoglobin. The mechanism by which the folding intermediate stabilizes specific α -helices is not yet known.

Additional support for the hypothesis of native-like secondary structures in folding intermediates comes from a study of an equilibrium folding intermediate of α -lactalbumin [12] and from peptide models of intermediates in disulfide bond formation of BPTI [13,14], whose structures are being analysed by direct NMR methods. One should be cautious, however, about generalizing the hypothesis of native-like structure in folding intermediates. A kinetic circular dichroism study of a β -lactoglobulin folding intermediate indicates that it has non-native secondary structure [15], a conclusion which is supported by the folding kinetics of another β -sandwich protein, fatty acid binding protein [16].

What are the prospects for characterizing the structures of kinetic folding intermediates? These cannot be studied by direct NMR methods, and exchange rates in kinetic intermediates cannot be measured in ambient folding conditions, as they can for equilibrium folding intermediates. Moreover, unfolded proteins are typically heterogeneous as regards their refolding behavior because of slow processes that occur in the unfolded protein, such as the slow *cis-trans* isomerization of prolyl peptide bonds or, in the case of cytochrome *c*, slow ligand exchange reactions with the heme iron.

Nevertheless, the kinetic folding experiment provides a simple test for structural folding intermediates: there is clear evidence of their presence if the folding kinetics measured by individual amide protons yield different curves. By this test, cytochrome *c*, ribonuclease A, and barnase all have structural intermediates. A recent study of the kinetics of folding of hen lysozyme by a competition method is also consistent with the presence of a partly folded intermediate [17]. Moreover, information about the protection factors of amide protons in kinetic folding intermediates has been obtained by pulse label experiments in which the pH of the pulse is varied [5,6].

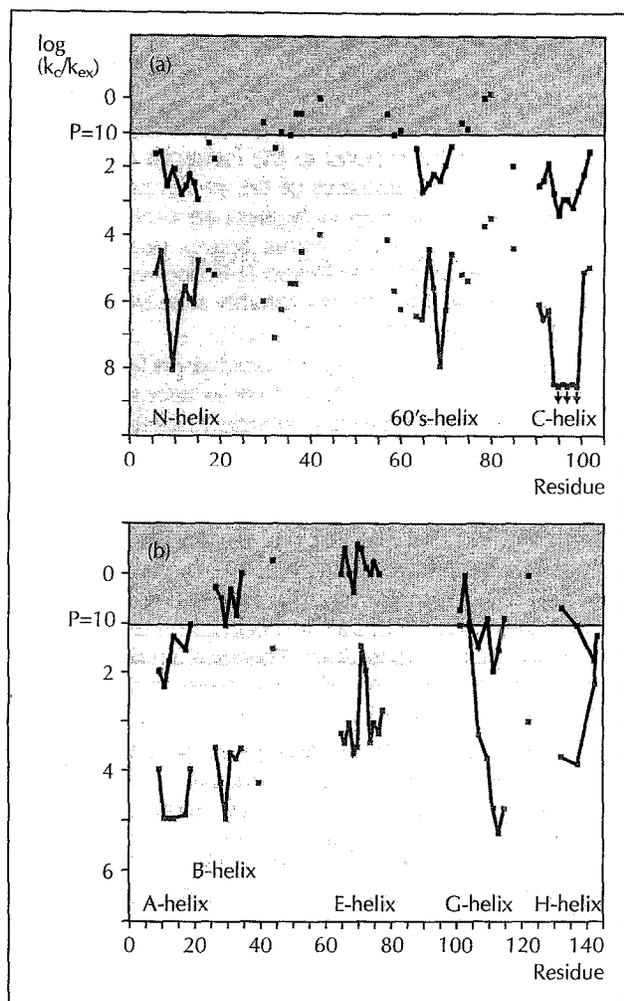


Fig. 2. The logarithm of the protection factor for amide proton exchange plotted against residue number (a) N (red) and I (blue) forms of horse cytochrome *c* and (b) N (red) and I (blue) forms of sperm whale apomyoglobin. Solid lines connect the residues that are α -helices in the N forms. Arrows mark residues whose exchange rates are too small to measure. (a) Adapted from [9].

When amide proton exchange is base-catalysed, as is typically the case, increasing the pH by one unit produces a 10-fold increase in the driving force for exchange. Not only are the N and C helices of cytochrome *c* formed with similar kinetics in an early folding intermediate [3], but the protection factors of amide protons in these two helices are similar in the folding intermediate [6].

The problem of obtaining clear-cut structural information about kinetic, rather than equilibrium, folding intermediates is more daunting. Two factors seem likely to help. First, the structures of equilibrium folding intermediates, which can be analysed both more directly and more readily, as discussed here, provide valuable guides of what to expect in kinetic experiments. Second, present information suggests that noncovalent folding reactions are hierarchical, in the sense that units of structure formed early in folding are retained throughout the folding process and are incorporated into the structure of the native protein. As already noted, it may prove nec-

essary to develop tests for abortive structure formation in order to make use of this principle. If the postulate of hierarchical folding proves to be correct, then pulsed amide proton exchange provides detailed structural information about the folding pathway. Comparison of the patterns of protection for a folding intermediate and the native protein can serve as a guide as to whether or not the folding process is hierarchical.

References

1. CREIGHTON TE: Protein folding. *Biochem J* 1990, 270:1-16.
2. UDGAONKAR JB, BALDWIN RL: NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 1988, 335:694-699.
3. RÖDER H, ELÖVE GA, ENGLANDER SW: Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* 1988, 335:700-704.
4. BYCROFT M, MATOUSCHEK A, KELLIS JT JR, SERRANO L, FERSHT AR: Detection and characterization of a folding intermediate in barnase by NMR. *Nature* 1990, 346:488-490.
5. UDGAONKAR JB, BALDWIN RL: Early folding intermediate of ribonuclease A. *Proc Natl Acad Sci USA* 1990, 87:8197-8201.
6. ELÖVE GA, RÖDER H: Structure and stability of cytochrome *c* folding intermediates. In *Protein Refolding*, edited by Georgiou G. Am Chem Soc Symposium Series, 470, in press.
7. CREIGHTON TE: Characterizing intermediates in protein folding. *Curr Biol* 1991, 1:8-10.
8. WAGNER GA, WÜTHRICH K: Amide proton exchange and surface conformation of BPTI in solution. *J Mol Biol* 1982, 160:343-361.
9. JENG M-F, ENGLANDER SW, ELÖVE GA, WAND AJ, RÖDER H: Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR. *Biochemistry* 1990, 29:10433-10437.
10. HUGHSON FM, WRIGHT PE, BALDWIN RL: Structural characterization of a partly folding apomyoglobin intermediate. *Science* 1990, 249:1544-1548.
11. JENG M-F, ENGLANDER SW: Stable submolecular folding units in a non-compact form of cytochrome *c*. *J Mol Biol* 1991, in press.
12. BAUM J, DOBSON CM, EVANS PA, HANLEY C: Characterization of a partly folded protein by NMR methods. Studies on the molten globule state of guinea pig α -lactalbumin. *Biochemistry* 1989, 28:7-13.
13. OAS TG, KIM PS: A peptide model of a protein folding intermediate. *Nature* 1988, 336:42-48.
14. STALEY JP, KIM PS: Role of a subdomain in the folding of bovine pancreatic trypsin inhibitor. *Nature* 1990, 344:685-688.
15. KUWAJIMA K, YAMAYA H, MIWA S, SUGAI S, NAGAMURA T: Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. *FEBS Lett* 221:115-118.
16. ROPSON IJ, GORDON JI, FRIEDEN C: Folding of a predominantly β -structure protein: rat intestinal fatty acid binding protein. *Biochemistry* 1990, 29:9591-9599.
17. MIRANKAR A, RADFORD SE, KARPLUS M, DOBSON CM: An NMR demonstration of folding domains in lysozyme. *Nature* 1991, 349:633-636.

Robert L. Baldwin, Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA. Heinrich Roder, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111, USA.