Finding Intermediates in Protein Folding

Robert L. Baldwin

Accepted 19 November 1993

In 1968 I was listening to a seminar at Stanford by Cy Levinthal entitled 'How to fold gracefully'. He proposed that some day it would be possible to predict the three-dimensional structures of proteins from their amino acid sequences by using computers to decide between enormous numbers of possible alternative conformations. Cy suggested that prediction of the final structure should be made step by step, at every stage of the folding process, and he pointed out that it was going to be very important to find out how proteins fold up. At that time the unfolding and refolding reactions of some small proteins were known to be freely reversible. Chris Anfinsen's conclusion[1] was widely accepted, that the driving force for the folding process is simple thermodynamics: a protein folds up by making side-chain interactions and peptide hydrogen bonds that stabilize its native structure. Two years earlier, in 1966, Lumry, Biltonen and Brandts had argued persuasively[2] that the reversible unfolding/refolding reactions of small proteins follow a two-state model (U⇔N, where U = unfolded, N = native) without observable intermediates. In his Stanford seminar, Cy Levinthal took their model and pushed it to a *reductio ad absurdum*. He made a simple calculation[3] showing that it would take longer than the lifetime of the universe for a small protein to fold up by a random search of all possible conformations. Paul Flory was sitting next to me at his seminar; he nudged me and whispered 'so there must be folding intermediates'.

The ideas suggested by Cy's seminar incubated in my mind and in 1970, when I was at a Gordon conference listening to a talk on the two-state model of folding (no detectable intermediates), I suddenly realized that I knew a possible sensitive method for detecting folding intermediates. Dietmar Pörschke and Manfred Eigen[4] had analyzed the kinetic process of forming the RNA double helix and had found that it follows a nucleation mechanism: first the double helix is stably nucleated by forming a few correct, adjacent base pairs and then the helix forms rapidly by zippering up the two complementary strands. The double helix unwinds easily from each end and both ends are frayed slightly open. The fraying reaction, in which the end base pairs open up, produces intermediates in unfolding and occurs in a much faster time range than the overall process either of helix formation or of complete unfolding. I reasoned that protein folding might be a similar nucleation-limited process and, if so, the last steps in the folding process might be easily reversed: theory indicated that this event should occur in a fast time range. We could test this model by putting a protein solution into our temperature jump apparatus which would use a fast (10^{-5} s) temperature jump to bring the solution into the temperature range where unfolding occurs.

I broached this idea to Tian Tsong, a newly arrived postdoctoral fellow from Julian Sturtevant's laboratory. Tian had made calorimetric studies of protein unfolding for his PhD thesis and he believed that unfolding intermediates should be detectable by the right approach. He promptly put the well-studied protein ribonuclease A (RNase A EC 3.1.27.5) into our Eigen-DeMaeyer T-jump apparatus and the experiment worked[5]! There was a significant fast reaction in unfolding in the 10 ms time range, more than 1000 times faster than the subsequent slow unfolding process. Eliot Elson's theoretical analysis[5] indicated that the rapid increase in extent of the fast reaction with temperature, inside the unfolding transition zone, could be explained by the nucleation-limited model of folding. We monitored the tyrosine absorbance change that occurs when previously buried tyrosine side chains are exposed to solvent. Only later did we learn that unfolding intermediates such as this one are rare beasts and that, if we had performed the experiment at neutral pH instead of using a low pH where RNase A unfolds at moderate temperatures, we would not have found an unfolding intermediate[6].

This experiment in 1971 led us to believe at first that nucleation-limited folding might be the right model for the folding process. Within a few months, however, Tian Tsong made a second experiment[7] showing that the folding of RNase A is not this simple. This time he used a stopped-flow apparatus to examine both the kinetics of refolding and unfolding. With a single pH jump, it is possible to pass through the entire transition zone either for unfolding or refolding, depending on the direction of the jump. He found a major fast reaction in refolding, indicative of a refolding intermediate[7]. But, in our nucleation-limited model of folding, although in principle there should be a nucleated refolding intermediate, in practice it should not be detectable, especially with the probe we were using (tyrosine absorbance, that monitors burial of tyrosine side chains). Two other groups independently found fast steps in the process of protein folding in 1971: Atsushi Ikai in Charles Tanford's laboratory[8] and Henry Epstein and coworkers[9] in Chris Anfinsen's laboratory.

As Tian Tsong left, a new postdoctoral fellow from Paris, Jean-Renaud Garel, took up the problem of deciphering the meaning of the fast reaction in refolding. We decided that most likely a folding intermediate was being formed in the fast folding reaction and, because we were monitoring its formation by using a probe for tertiary structure (burial of tyrosine side chains), the intermediate might be sufficiently folded to bind 2′CMP, a competitive inhibitor of RNase A. There is a nicely measurable change in absorbance when 2′CMP binds to native RNase A. We were surprised but gratified when our experiments showed that the putative folding intermediate, formed in the fast refolding reaction, does bind 2′CMP[10]. We then set out to compare other properties of the folding intermediate, such as its affinity for 2′CMP and specific tyrosine absorbance, to those of the native protein. These proved to be indistinguishable from the corresponding

properties of native RNase A, and we were forced to conclude that the product of the fast refolding reaction was, in fact, the native protein[10], not a folding intermediate. Our attention then became focused on possible heterogeneity in the unfolded protein: perhaps the starting material for the fast refolding reaction was incompletely unfolded, and thus was already nucleated for refolding? After using high temperatures to melt out any residual structure, we were also forced to conclude that the unfolded protein was completely unfolded[10]. A later study, using the denaturant 6 M GdmCl (guanidinium chloride), confirmed this conclusion[11].

Thus, we had two separate refolding reactions of RNase A, one 100 times faster than the other, but both reactions began with completely unfolded protein and both reactions yielded native protein. We concluded[10] that we must have two separate unfolded forms, a slow-folding form U_S and a fast-folding form U_F, always in a definite 20:80 U_F:U_S ratio, but we didn't know the origin of these two different forms. Paul Hagerman undertook a thorough experimental and theoretical analysis[12] of the $U_S \Leftrightarrow U_F \rightarrow N$ mechanism for the kinetics of unfolding and refolding and found that this mechanism accurately predicted the results, after the presence of Tian Tsong's unfolding intermediate[5] was taken into account.

In 1975 John Brandts and coworkers[13] proposed a plausible solution to the paradox: they suggested that the U_S form of RNase A arises from the U_F form after unfolding by *cis* ⇔ *trans* isomerization of the proline residues. Isomerization about a proline peptide bond was known to be a slow reaction, as is *cis* ⇔ *trans* isomerization of ordinary peptide bonds but, unlike other peptide bonds, the *cis* isomer of a prolyl peptide bond occurs fairly commonly in native proteins and there are two *cis* prolyl peptide bonds in native RNase A. After unfolding, a *cis* prolyl peptide bond is expected to isomerize chiefly to the *trans* form (10:90 or 20:80 are typical *cis*:*trans* ratios for isomers of prolyl peptide bonds) and, if a single wrong prolyl isomer can block the folding of a small protein, even a single *cis* isomer in the native protein might account for a 20:80 U_F:U_S ratio in the unfolded protein.

The proline model raised two problems for us: the first was to find out if it gave the correct explanation for the fast- and slow-folding forms of RNase A, and the second was to find out, if so, whether proline isomerization was the whole story, or whether instead the folding kinetics also revealed the presence of structural folding intermediates. I was optimistic about the answer to the second question: it seemed unlikely that proline isomerization could explain the unfolding intermediate we had found initially[5].

Barry Nall and Jean-Renaud Garel[14] undertook a study of whether the kinetics of heat and GdmCl-induced unfolding and refolding of RNase A were consistent with the proline model. Proline isomerization in unstructured peptides has some characteristic kinetic properties, notably a very large activation enthalpy (about 20 kcal/mol) and a rate that is nearly independent of denaturant concentration. They found a perplexing picture[14]: some properties of the unfolding and refolding kinetics of RNase A are consistent with the proline model, but others indicate that structural folding intermediates are needed to explain the results. Franz Schmid decided to make a test for a kinetic property that, if found, could only be ascribed to isomerization of prolyl or other peptide bonds: catalysis of *cis-trans* isomerization in the unfolded protein by strong acid. I. Z. Steinberg, Bill Harrington and their coworkers[15] had shown that proline *cis-trans* isomerization in poly-L-proline, dissolved in organic solvents, is catalyzed by strong acid. Franz tested for acid catalysis of the $U_F \Leftrightarrow U_S$ reaction of unfolded RNase A in aqueous solution. Significant catalysis was observed only above 3 M $HClO_4$, but there was a 100-fold increase in rate at 10 M $HClO_4$, and 80% native RNase A could be recovered after exposure to 10 M $HClO_4$ at 0°C[16]. His results left no doubt that proline *cis-trans* isomerization is responsible for the $U_F \Leftrightarrow U_S$ reaction in unfolded RNase A.

Franz Schmid's work[16] put us in a position to make a direct test for structural folding intermediates in refolding. At 0° or 10°C the rate of proline isomerization is exceedingly slow, and we could ask if the presence of one or two wrong proline isomers in a U_S molecule is able to block folding completely at early times, or whether partial folding occurs rapidly and precedes proline isomerization. Kem Cook and Franz Schmid found that partial folding of U_S does occur rapidly at 10°C, even when monitored by a probe for tertiary structure, the burial of tyrosine side chains[17]. It was possible to demonstrate that proline isomerization had not yet occurred when the intermediate was formed because, when the partly folded molecules were unfolded they were found to be U_S, meaning that the refolding intermediate still contained one or more wrong proline isomers. Once folding goes to completion and native RNase A is formed, unfolding gives U_F as the directly formed product. Further study of the partly folded intermediate[17] showed that it was surprisingly native-like: it could bind 2′CMP and contained buried tyrosine side-chains. We named this folding intermediate the native-like intermediate (I_N), and were quite surprised to find that isomerization of the wrong proline isomer(s) proceeded more rapidly in I_N than in unfolded RNase A! We had supposed that formation of tertiary structure would cause proline isomerization inside a folded protein to become exceedingly slow.

Having found clear evidence for a structural folding intermediate, we began to feel confident about the next steps. The first question was what probes to use to test for earlier structural intermediates in folding. During a sabbatical leave spent with Kai Linderstrøm-Lang at the Carlsberg Laboratory in Copenhagen in 1958-59, I became very interested in Lang's vision of using hydrogen isotope exchange between peptide NH protons and solvent to monitor the presence of α-helices and β-sheets. Peptide hydrogen exchange is inhibited by hydrogen-bonded structure, because an existing peptide H-bond must break before exchange can occur, and cooperatively H-bonded structures such as helices and sheets were expected to be particularly resistant to exchange. Franz Schmid[18] tested for early kinetic intermediates that resist hydrogen exchange during the folding of RNase A. At the start of folding, he set up a competition between exchange out of ^3H label from peptide NH protons and retention of label that must result from the presence of early folding intermediates. He found that early folding intermediates that

resist exchange are present and we concluded, cautiously, that at least some secondary structure is formed early during the folding of RNase A, before any intermediates are present with tertiary structures and buried tyrosine side chains. A year later Peter Kim worked out a pulse-labeling method[19] for measuring H-bonded folding intermediates in a simpler manner, and his results confirmed those of Franz Schmid.

We had to wait eight years to adapt the pulse-labeling experiment to the study of individual peptide NH protons, by using 2-D proton NMR to analyze the pulse label results. First, we had to obtain adequate access to a high-resolution NMR instrument and then we had to make the resonance assignments of RNase A. Andy Robertson made the resonance assignments[20], in collaboration with Enrico Purisima and Margaret Eastman from Harold Scheraga's laboratory, and Jayant Udgaonkar[21] made our first pulsed-hydrogen exchange experiments, analyzed by 2-D NMR, on the folding pathway of RNase A. At the University of Pennsylvania, Heinrich Roder and Walter Englander and their coworkers made a parallel study of the folding pathway of cytochrome *c*[22]. We knew about each other's work and exchanged information about these first experiments, and arranged to have our two papers published side-by-side. The locations of peptide NH protons used as probes in this work are shown in Fig. 1.

This approach to the problem of determining folding pathways caught on rapidly, and a recent review[23] summarized pulsed-hydrogen exchange results from eight laboratories on the folding pathways of nine different proteins. The remarkable fact is that, after many years in which most workers thought that structural folding intermediates could not be detected, almost every protein that has been studied by this approach shows one or more well-populated intermediates during the kinetic process of folding. Progress in analyzing folding pathways is rapid, and workers in this field are close to agreement on the overall features of refolding pathways.

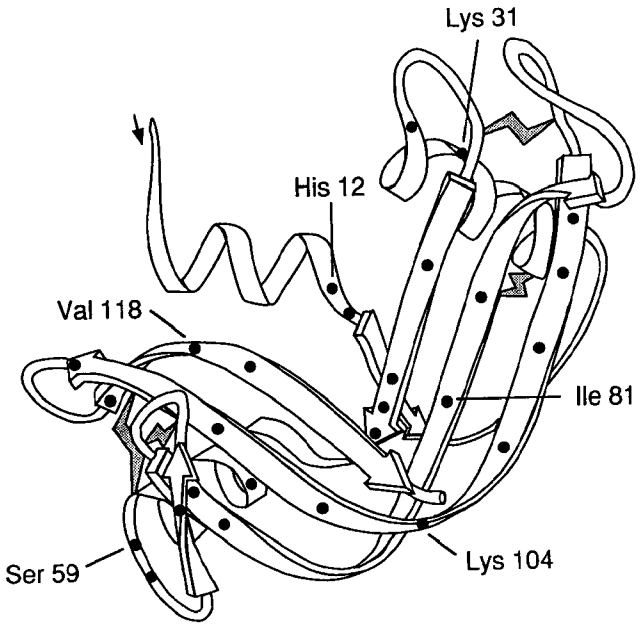

Fig. 1. Ribbon diagram of the structure of ribonuclease A (diagram courtesy of Jane Richardson) showing the locations of peptide NH protons used as probes of the kinetic pathway of refolding[34] (reprinted, with permission). The two *cis* proline residues of RNase A are in loops at opposite ends of the molecule: Pro 93 is in the unmarked loop at the upper right and Pro 114 is in the loop at the lower left, close to Val 118.

Our kinetic approach to finding folding intermediates was one of three approaches originating in the early 1970s that led to the analysis of folding pathways. In 1974, Tom Creighton demonstrated covalent trapping of kinetic intermediates in the linked pathways of disulphide bond formation and folding[24]. The folded structures of proteins that contain S-S bonds are stabilized by these bonds, so that S-S intermediates may be considered as folding intermediates, and S-S intermediates can be trapped by blocking unreacted -SH groups with iodoacetate. The third approach to finding folding intermediates resulted from the accidental discovery that a few rare proteins show partly folded structures at low pH, or in moderate concentrations of urea or guanidinium chloride. These equilibrium folding intermediates were discovered in 1973 for carbonic anhydrase[25] and β-lactamase[26], and a few years later for α-lactalbumin[27], which came to be regarded as the paradigm of a molten globule intermediate[28-30]. Even though they are populated at equilibrium in highly non-physiological conditions, molten globule intermediates are recognized today as showing characteristic properties of early intermediates in the kinetic process of folding[23,29,30,31]. In the late 1980s, a fourth, very powerful, approach to analyzing folding intermediates came into being: the design and synthesis of peptide models for folding intermediates[32].

The role of chaperones in assisting the process of folding *in vivo* is a separate and fascinating story that has developed rapidly in the last few years[33]. Most workers expect to find close connections between *in vitro* folding intermediates and the protein conformations that are stabilized by binding to chaperones; we should know fairly soon if this expectation is correct.

Acknowledgement

This research has been supported by a grant (GM 19988) from the Institute of General Medical Sciences, National Institutes of Health.

References

1 Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

2 Lumry, R., Biltonen, R. and Brandts, J.F. (1966). Validity of the 'two-state' hypothesis for conformational transitions of proteins. *Biopolymers* **4**, 917-944.

3 Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Physique* **65**, 44-45.

4 Pörschke, D. and Eigen, M. (1971). Cooperative non-enzymatic base recognition. III. Kinetics of the helix-coil transition of the oligoribouridylic-oligoriboadenylic acid system and of oligoriboadenylic acid alone at acid pH. *J. Mol. Biol.* **62**, 361-381.

5 Tsong, T.Y., Baldwin, R.L. and Elson, E.L. (1971). The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc. Natl Acad. Sci. USA* **68**, 2712-2715.

6 Tsong, T.Y. and Baldwin, R.L. (1978). Effects of solvent viscosity and different guanidine salts on the kinetics of ribonuclease A chain folding. *Biopolymers* **17**, 1669-1678.

7 Tsong, T.Y., Baldwin, R.L. and Elson, E.L. (1972). Properties of the refolding and unfolding reactions of ribonuclease A. *Proc. Natl Acad. Sci. USA* **69**, 1809-1812.
8 Ikai, A. and Tanford, C. (1971). Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* **230**, 100-102.
9 Epstein, H.F., Schechter, A.N., Chen, R.F. and Anfinsen, C.B. (1971). Folding of staphylococcal nuclease: kinetic studies of two processes in acid renaturation. *J. Mol. Biol.* **60**, 499-508.
10 Garel, J.-R. and Baldwin, R.L. (1973). Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc. Natl Acad. Sci. USA* **70**, 3347-3351.
11 Garel, J.-R., Nall, B.T. and Baldwin, R.L. (1976). Guanidine-unfolded state of ribonuclease A contains both fast- and slow-refolding species. *Proc. Natl Acad. Sci. USA* **73**, 1853-1857.
12 Hagerman, P.J. and Baldwin, R.L. (1976). A quantitative treatment of the folding transition of ribonuclease A. *Biochemistry* **15**, 1462-1472.
13 Brandts, J.F., Halvorson, H.R. and Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry* **14**, 4953-4963.
14 Nall, B.T., Garel, J.-R. and Baldwin, R.L. (1978). Test of the extended two-state model for the kinetic intermediates observed in the folding transition of ribonuclease A. *J. Mol. Biol.* **118**, 317-330.
15 Steinberg, I.Z., Harrington, W.F., Berger, A., Sela, M. and Katchalski, E. (1960). The configurational changes of poly-L-proline in solution. *J. Am. Chem. Soc.* **82**, 5263-5279.
16 Schmid, F.X. and Baldwin, R.L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA* **75**, 4764-4768.
17 Cook, K.H., Schmid, F.X. and Baldwin, R.L. (1979). Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proc. Natl Acad. Sci. USA* **76**, 6157-6161.
18 Schmid, F.X. and Baldwin, R.L. (1978). Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. *J. Mol. Biol.* **135**, 199-215.
19 Kim, P.S. and Baldwin, R.L. (1980). Structural intermediates trapped during the folding of ribonuclease A by amide proton exchange. *Biochemistry* **19**, 6124-6129.
20 Robertson, A.D., Purisima, E.O., Eastman, M.A. and Scheraga, H.A. (1989). Proton NMR assignments and regular backbone structure of bovine pancreatic ribonuclease A in aqueous solution. *Biochemistry* **28**, 5930-5938.
21 Udgaonkar, J.B. and Baldwin, R.L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694-699.
22 Roder, H., Elöve, G.A. and Englander, S.W. (1988). Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* **335**, 700-704.
23 Baldwin, R.L. (1993). Pulsed H/D-exchange studies of folding intermediates. *Curr. Opin. Struct. Biol.* **3**, 84-91.
24 Creighton, T.E. (1974). Intermediates in the refolding of reduced pancreatic trypsin inhibitor. *J. Mol. Biol.* **87**, 579-602.
25 Wong, K.-P. and Tanford, C. (1973). Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride: a process involving separable sequential conformational transitions. *J. Biol. Chem.* **248**, 8518-8523.
26 Robson, B. and Pain, R.H. (1973). Studies on the role of α-helix in nucleation. In *Conformation of Biological Molecules and Polymers*, vol. 5 (ed. E.D. Bergmann and B. Pullman), pp. 161-172. The Jerusalem Symposia on Quantum Chemistry and Biochemistry.
27 Kuwajima, K. (1977). A folding model of α-lactalbumin deduced from the three-state denaturation mechanism. *J. Mol. Biol.* **114**, 241-258.
28 Dolgikh, D.A., Gilmanshin, R.I., Brazhnikov, E.V., Bychkova, U.E., Semisotnov, G.V., Venyaminov, S.Y. and Ptitsyn, O.B. (1981). α-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett.* **136**, 311-315.
29 Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Struct. Funct. Genet.* **6**, 87-103.
30 Ptitsyn, O.B. (1992). The molten globule state. In *Protein Folding* (ed. T.E. Creighton), pp. 243-300. Freeman and Co., New York.
31 Jennings, P.A. and Wright, P.E. (1993). Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-896.
32 Oas, T.G. and Kim, P.S. (1988). A peptide model of a protein folding intermediate. *Nature* **336**, 42-28.
33 Jaenicke, R. (1993). Role of accessory proteins in protein folding. *Curr. Opin. Struct. Biol.* **3**, 104-112.
34 Udgaonkar, J.B. and Baldwin, R.L. (1990). Early folding intermediate of ribonuclease A. *Proc. Natl Acad. Sci. USA* **87**, 8197-8201.

Robert L. Baldwin is at the Department of Biochemistry, School of Medicine, Beckman Center, Stanford University, Stanford, CA 94305-5307, USA.