Structural studies of protein unfolding

Girish S. Ratnaparkhi and R. Varadarajan

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Methods for macromolecular structure determination (NMR and crystallography) are now being used to get structural information on partially folded and unfolded states of proteins. These techniques, in combination with proton hydrogen exchange studies are powerful tools to extract information on non-native states of proteins. This review discusses progress in this area of protein folding.

PROTEINS are biopolymers of twenty naturally occurring amino acids. Protein chains are typically between 50 and 500 amino acids in length. Each protein is characterized by one (or in a few cases, more than one) folded chain conformation. An important unsolved biological problem, called the 'protein folding problem'[1–3], is understanding the relationship between the amino acid sequence of a protein and its folded three-dimensional structure. At present (and in the foreseeable future) the number of known protein sequences is far in excess of the number of known protein structures. Protein sequences can be readily determined from the DNA sequence of the corresponding gene. In contrast, protein structures can be determined to atomic resolution only by X-ray crystallography or nuclear magnetic resonance (NMR) techniques. Both these techniques are time consuming, technically difficult and require large amounts (10–100 mg) of protein. Since protein structure is intimately linked with biological function it would be highly desirable to predict structure from the amino acid sequence.

Protein folding is a search for the final folded state starting from an unfolded, ensemble of conformationally flexible states. A folded protein chain can theoretically adopt a very large number of sterically allowed conformations. However, many proteins are able to fold from an unfolded random coil state to the final folded state on timescales ranging from milliseconds to minutes. In these timescales only a tiny fraction of all sterically allowed conformations can be sampled. Protein folding is often well described by a two-state process in which only the folded state N (with well-defined three-dimensional structure) and the unfolded state U (which consists of an ensemble of interconverting conformations) are significantly populated. An understanding of the pathway(s) by which a given protein folds might be useful in the solution of the protein folding problem[2,3]. However, because the protein folding process is rapid and often takes place without detectable intermediates (states different in structure or energetics from both the folded and the unfolded state), this is a challenging task. Much effort has gone into identifying solution conditions (pH, temperature, denaturant or additive concentration) under which folding intermediates are populated to a significant extent, either kinetically or at equilibrium[4,5]. If conditions can be identified under which folding intermediates are in equilibrium with either the folded or the unfolded state, then the conformation of such states can be probed by a variety of spectroscopic and hydrodynamic techniques. In such cases it is important to show that such equilibrium intermediates are similar to intermediates formed along the kinetic pathway of folding of the protein[6]. It is also of interest to characterize the thermodynamic parameters (especially the free energy changes) involved in transitions from unfolded and partially folded states to the folded state. As summarized below, denaturants such as urea, guanidine hydrochloride (GdnCl), pH or temperature are used both in kinetic and thermodynamic studies of protein folding[2,7].

In order to measure the free energy of folding of a protein (ΔG^0), it is necessary to find a range of conditions under which both the folded state (N) and the unfolded state (U) are populated to significant extents. The equilibrium constant $K_f = [N]/[U]$ can then be accurately determined as a function of denaturant concentration or temperature. However, the values of ΔG^0 ($\Delta G^0 = -RT \ln K_f$) measured under these conditions need to be extrapolated to zero denaturant concentration, neutral pH and room temperature. In the case of urea and guanidine hydrochloride denaturation, a linear extrapolation procedure is widely used, $\Delta G^0_D = \Delta G^0 + m^*[D]$, where ΔG^0_D is the free energy of folding at a denaturant concentration [D] and m is the slope of a plot of ΔG^0_D vs [D]. Despite their extensive use, the structural basis of denaturation by urea and GdnCl is still not clear[8,9]. Structural information has been obtained indirectly by biophysical studies like spectroscopy[10] and thermodynamics[11] .There are two mechanisms by which urea is thought to destabilize proteins. Urea may interact with backbone peptide groups through the formation of hydrogen bonds. The crystal structure of the diketopiperazine–urea complex[12] shows that the diketopiperazine molecule forms several hydrogen bonds with urea molecules. Solvation of the diketopiperazine by urea is so extensive that there are no hydrogen bonds between diketopiperazine molecules in the crystal. However it should be noted that this

structural information is derived from a tightly packed small molecule crystal comprising solely of urea and diketopiperazine. The interactions of urea with the peptide backbone in aqueous solution where a large amount of aqueous hydrogen bonding solvent is present, may be quite different. Another proposed mechanism for urea denaturation is that high concentrations of urea decrease the magnitude of the hydrophobic driving force. Support for this comes from solubility data of hydrocarbons in urea[13]. Thermodynamic studies of the binding of urea to proteins and peptides[11] suggest that urea binds weakly to the protein with dissociation constants in the molar range. A study of the specificity of urea binding to the folded states of two small proteins, BPTI and PEC-60, was recently carried out by monitoring the NMR chemical shifts[14] of protein protons as a function of urea concentration. Binding constants obtained by this technique were similar to those obtained calorimetrically[11]. However, neither of these two proteins is denatured by urea. In order to clarify the mechanism of urea denaturation, it is important to carry out structural studies of proteins and peptides that undergo denaturation in the presence of urea. Although NMR and X-ray crystallography are the two most popular tools for macromolecular structure determination, there are problems with application of these techniques to determining the structure of unfolded and partially folded proteins. Such states generally do not show well resolved NMR peaks and display much smaller dispersion in chemical shifts than in the folded state. No partially unfolded state of a protein has been crystallized to date, probably because disordered regions of the protein act as an impediment to crystallization. Partially unfolded states also show a tendency to aggregate at the high concentrations required for crystallography and NMR. However, with increases in the magnetic field strengths available as well as with the advent of heteronuclear NMR experiments it has been possible to assign spectra of proteins in the unfolded state where there is extensive resonance overlap due to conformational averaging[15–17]. This is the first step in structural characterization of the unfolded state though much progress remains to be made. NMR has also been used to provide indirect structural information on unfolded and partially folded states through hydrogen exchange studies.

One of the most informative techniques used in studying the process of protein folding is that of hydrogen exchange[18–22]. Protons in a protein that are bonded to nitrogen or oxygen atoms can exchange with protons from the aqueous solvent. The rate at which exchange occurs is determined by several factors such as solution pH and temperature, the accessibility of a given proton to solvent, whether the proton is hydrogen bonded to another acceptor in the protein and the conformational stability of the protein. For many protons on the surface of the protein, exchange is a rapid process that occurs in less than a second. Interior protons have much slower exchange rates and in some cases half lives of exchange can be several months. A commonly used model for describing hydrogen exchange is as follows. The hydrogen exchange rate constant of an amide proton that is fully accessible to solvent is given by,

$$k_3 = k_a [H^+] + k_b [OH^-] + k_w \quad (1)$$

For a protected amide proton exchange can be modelled as shown below,

$$C \underset{k_2}{\overset{k_1}{\Leftrightarrow}} O \overset{k_3}{\Rightarrow} EX \quad (2)$$

Here C and O are the closed and open states for a given amide proton. Exchange is assumed to take place only from the open state with a rate constant of k_3. There are two limiting cases for this model. In the first case, the so-called EX1 limit, $k_2 \ll k_3$. Then the overall rate constant for exchange k_{ex} is equal to k_1, the rate constant for opening of the protein to an exchange competent state. In the second limiting case, the so-called EX2 limit, $k_3 \ll k_2$. Hence $k_{ex} = (k_1/k_2)*k_3 = K*k_3$, where K is the equilibrium constant for the transition between C and O. Exchange rate constants for exchange that occurs by the EX1 mechanism will be relatively independent of pH. In contrast, for the EX2 case, the exchange rate constant is expected to depend on pH according to equation (1). In the majority of cases, exchange appears to take place by the EX2 mechanism though cases of EX1 exchange have also been observed[22,23]. Values of k_3 for a given amide proton can be estimated from exchange rate measurements in small peptides[24]. Thus measurements of k_{ex} for EX2 exchange can be used to determine the equilibrium constant K for unfolding to an exchange competent state. Much recent work has focused on measurements of exchange rates of the native state (and hence K and ΔG^0 and m for the C to O transition) as a function of denaturant concentration[23,25–27]. These exchange rates for individual amide protons can be made using 2D NMR. Proteins studied include RNase H (ref. 28), RNase A (ref. 29) and Cyt c (ref. 30). From such measurements it is possible to determine whether an exchange event for a given proton is coupled to global or local unfolding. ΔG^0 and m values for individual amide protons determined from such exchange measurements can be compared with the ΔG^0 and m values obtained from spectroscopic measurements of global protein folding. If the two values are identical, this suggests that the O state for the amide proton is identical to the globally unfolded state. This method of obtaining information about folding intermediates and local unfolding is termed

'native state' exchange because studies are carried out at low denaturant concentrations in which most protein molecules are in the native state.

Another hydrogen exchange technique that has provided much useful information is the use of NMR to determine at what stage during the folding pathway a specific amide proton becomes protected from exchange[29–32]. As a protein folds, some amide protons become inaccessible to further exchange. By appropriately changing the pH during the course of folding, it is possible to restrict exchange to occur only at defined periods during the folding process. The NMR spectrum can be acquired once the molecule has folded. The extent of exchange occurring during the exchange period can therefore be quantified. From this information it is possible to obtain a qualitative idea of the structure present in the protein during the time interval that exchange occurred[31,32]. In cases where equilibrium intermediates have been identified, hydrogen exchange studies[6,20,33] have also been used to provide indirect structural information about such molecules.

X-ray crystallography is widely used to determine macromolecular structure. However it has not been widely used to study the protein folding process or to examine the effects of denaturants on protein structure. This technique provides information complementary to that obtained from NMR. Unlike NMR, which is currently limited to proteins of molecular weight less than 30 kD, crystallography has no size restriction. While NMR hydrogen exchange studies monitor primarily the dynamic behaviour of amide protons, in crystallography it is, in principle, possible to obtain the structure of the protein at atomic resolution as a function of denaturant concentration. While NMR can also be used for protein structure determination, the positional accuracy of coordinates obtained by this technique is still not comparable to that of a high resolution crystal structure.

Crystallographic studies in the presence of denaturants are limited to the study of a peptide model in urea[12], high resolution crystal structures of hen egg white lysozyme in the presence of denaturant concentrations as high as 8 M (ref. 34) and a low resolution structure of alpha chymotrypsin in the presence of GuHCl and urea[35]. In the case of lysozyme little change in protein structure was observed, probably because lysozyme is stable to urea denaturation at room temperature and physiological pH. With an increase in urea concentration, an increase in the order of the protein chain was observed in the loops and a few ordered urea molecules could be seen in the structure. Many of these occurred at bridging positions between symmetry-related molecules in the crystal. In the case of α-chymotrypsin, the protein undergoes denaturation in the presence of urea and it was not possible to collect high resolution X-ray data. This study was performed several years ago. Since then, data collection and refinement techniques have improved considerably. It is now possible to routinely collect datasets in a day, instead of in several weeks as was the case previously.

We have recently determined the structure of the protein–peptide complex RNase S in different concentrations of urea (Ratnaparkhi and Varadarajan, unpublished results). The objective of these studies is to obtain information on the structural changes associated with urea denaturation of this molecule. RNase S consists of S peptide (residues 1–20) and S protein (residues 21–124), two proteolytic fragments of the 124 amino acid protein, RNase A (Figure 1). These fragments are obtained by cleavage of the peptide bond between Ala-20 and Ser-21 of RNase A by subtilisin. The two fragments can be reconstituted to give a non-covalently bound, active, S peptide, S protein complex, RNase S (ref. 36). The residues 16–20 of the S peptide are not involved in binding of S peptide to S protein. The refined

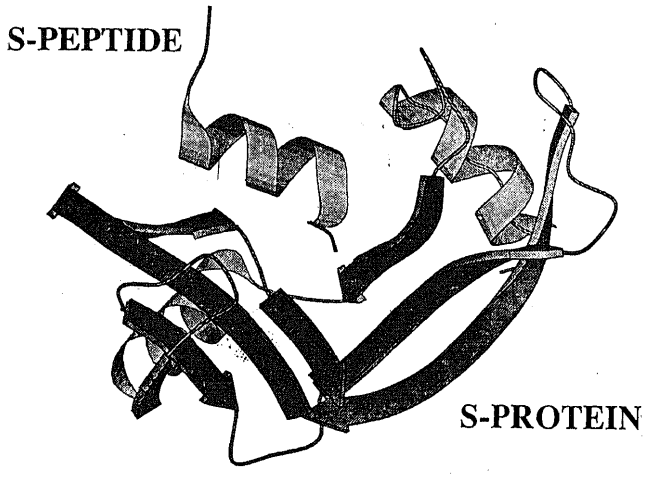

Figure 1. Ribbon diagram of ribonuclease S (ref. 38) drawn using the program MOLSCRIPT.

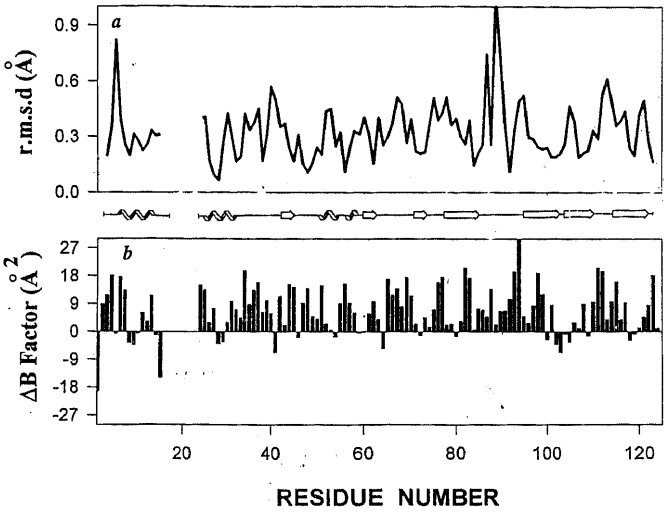

Figure 2. Plots representing the structure of RNase S in 5 M urea. The secondary structure along the sequence is represented as follows; Helix { }, beta-sheet { }, loops {—}. The break in the graph is because of lack of electron density for residues 16–23. *a*, Root mean square derivation plot; *b*, ΔB factor plot.

three-dimensional structures of RNase A and RNase S are now available to a high resolution[37,38] and the two proteins have very similar structures. We have previously characterized the thermodynamic and structural changes associated with a large number of residue substitutions in RNase S (refs 39–42). Thermodynamics of binding of various S-peptide analogues to S-protein were measured by titration calorimetry and structures of the bound complexes were characterized by X-ray crystallography. A significant advantage of the RNase S system is that it is a bimolecular system and hence effects of mutations on molecular stability can be studied by isothermal titration calorimetry[39]. In contrast, for most proteins, protein folding is a unimolecular reaction. Hence mutational effects on stability are typically studied using either spectroscopic or differential scanning calorimetric measurements. Measurements of important parameters such as the ΔH^0, ΔG^0 and ΔC_p of folding made by titration calorimetry are considerably more accurate than measurements made using either differential scanning calorimetry or spectroscopy. RNase S is readily crystallized and crystals typically diffract to a resolution of better than 1.8 Å. In solution at pH 8, RNase S is 50% denatured at a urea concentration of 3.5 M. We have soaked crystals of RNase S in a range of urea concentrations (0 M, 1.5 M, 2 M, 3 M and 5 M) and have solved these structures to resolutions ranging from 1.8 to 3.0 Å. At urea concentrations greater than 2 M, crystals were stabilized by crosslinking with glutaraldehyde before soaking in urea. In each case, a control data set was collected on a crystal crosslinked for an identical period of time but which was not soaked in urea. Crosslinking for up to 30 min had no effect on either the resolution of the data or the three-dimensional structure of the molecule. At higher crosslinking times (greater than 40 min) there was a rapid decrease in the resolution to which the crystals diffracted. In the urea-soaked crystals, the largest changes were observed in the structure of the complex at 5 M urea. Shown in Figure 2 *a* is the root mean square difference in coordinates of the main chain coordinates of the 5 M urea structure relative to the control structure in the absence of urea. These differences are plotted as a function of residue number. Figure 2 *b* shows the difference in the average main chain B factor of the 5 M urea structure relative to its control. Also indicated at the top of the figure are the positions of α helices, β strands and loops/turns in the sequence. There are significant changes in both the main chain coordinates and in the B-factors relative to the control structure. These differences are largest for the loop regions, especially for the loop between residues 65 and 72. Interestingly, there are also large changes observed in the position of this loop in mutants of RNase S which are significantly destabilized with respect to the wild type complex[41], suggesting that movement of this loop may be an initial step in the unfolding of the complex. It is also of interest to note that the β strand regions on average show smaller increases in B-factor than the α helical regions. These results are consistent with earlier exchange studies of RNase A (ref. 31) which show that the β sheet appears to be protected at earlier times in folding than any of the α helices. These results need to be substantiated by solving other RNase S crystal structures as a function of denaturant concentrations in different crystal forms. X-ray data for the structures discussed above were

collected at room temperature using a rotating anode X-ray source. Recent advances in X-ray crystallography include the use of more powerful X-ray sources, flash freezing techniques to reduce radiation decay as well as techniques such as multiwavelength anomalous dispersion (MAD) phasing to obtain a more accurate image of the structure and solvent without calculation of phases[43] from a model. Our work demonstrates that it is possible to use crystallography to solve protein crystal structures in the presence of denaturants and we anticipate that such structures will provide a wealth of information about the structural basis for denaturation by different denaturants as well as on partially folded states of proteins.

1. Richards, F. M., *Sci. Am.*, 1991, **264**, 54–57.
2. Utpal, N. and Udgaonkar J. B., *Curr. Sci.*, 1997, **72**, 180–190.
3. Ptitsyn, O. G., *FASEB J.*, 1996, **10**, 3–4.
4. Kim, P. S. and Baldwin, R. L., *Annu. Rev. Biochem.*, 1990, **59**, 631–660.
5. Creighton, T. E., Darby, N. J. and Kemmink, J., *FASEB J.*, 1996, **10**, 110–118.
6. Buck, M., Radford, S. E. and Dobson, C. M., *Biochemistry*, 1993, **32**, 669–678.
7. Pace, C. N., *CRC Crit. Rev. Biochem.*, 1975, **3**, 1–4.
8. Schellman, J. A., *Biophys. Chem.*, 1990, **37**, 121–140.
9. Prakash, V., Loucheux, C., Scheufele, S., Gorbunoff, M. J. and Timasheff, S., *Arch. Biochem. Biophys.*, 1981, **210**, 455–464.
10. Eftink, M. R., *Methods Enzymol.*, 1995, **259**, 487–512.
11. Makhatadze, G. I and Privalov, P. L., *J. Mol. Biol.*, 1992, **224**, 715–23.
12. Thayer, M. M., Haltiwanger, R. C., Allured, V. S., Gill, S. C. and Gill, S. J., *Biophys. Chem.*, 1993, **46**, 165–169.
13. Creighton, T. E., *Proteins*, W. H. Freeman and Co., New York, 1993, pp. 292–296.
14. Liepinsh, E. and Otting, G., *J. Am. Chem. Soc.*, 1994, **116**, 9670–9674.
15. Logan, T. M., Theriault, Y. and Fesik, S. W., *J. Mol. Biol.*, 1994, **236**, 637–648.
16. Zhang, O. and Forman-Kay, J. D., *Biochemistry*, 1995, **34**, 6784–6794.
17. Pan, H., Barbar, E., Barany, G. and Woodward, C., *Biochemistry*, 1995, **34**, 13974–13981.
18. Woodward, C., Simon, I. and Tuchsen, E., *Mol. Cell. Biochem.*, 1982, **48**, 135–160.
19. Englander, S. W., Kallenbach, N. R., *Q. Rev. Biophys.*, 1984, **16**, 521–655.
20. Shortle, D. R., *Curr. Op. Struct. Biol.*, 1996, **6**, 24–30.
21. Englander, S. W., Sosnick, T. R., Englander, J. J. and Leland, M., *Curr. Op. Struct. Biol.*, 1996, **6**, 18–23.
22. Kiefhaber, T. and Baldwin, R. L., *Biophys. Chem.*, 1996, **59**, 351–356.
23. Mayo, S. L. and Baldwin, R. L., *Science*, 1993, **262**, 873–876.
24. Bai, Y., Milne, J. S., Mayne, L. M. and Englander, S. W., *Prot. S.F. & G.*, 1993, **17**, 75–86.
25. Woodward, C. K. and Rosenberg, A., *J. Biol. Chem.*, 1971, **240**, 4114–4121.
26. Quian, H., Mayo, S. L. and Morton, A., *Biochemistry*, 1994, **33**, 8167–8171.
27. Kim, K. S., Woodward, C., *Biochemistry*, 1993, **32**, 9609–9613.
28. Chamberlain, A. K., Handel, T. N. and Marqusee, S., *Nat. Str. Biol.*, 1996, **3**, 782–787.
29. Udgaonkar, J. B. and Baldwin, R. L., *Nature*, 1988, **335**, 694–699.
30. Roder, H., Elove, G. A. and Englander, S. W., *Nature*, 1988, **355**, 700–704.
31. Baldwin, R. L., *Curr. Op. Struct. Biol.*, 1993, **3**, 84–91.
32. Englander, S. W. and Mayne, L., *Annu. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 243–265.
33. Haezebrouck, P., Joniau, M., Dael, H. M., Hooke, S. D., Woodruff, N. D. and Dobson, C. M., *J. Mol. Biol.*, 1995, **246**, 382–387.
34. Pike, C. W. and Acharya, R., *Prot. Sci.*, 1994, **3**, 706–710.
35. Hibbard, L. S. and Tulinsky, A., *Biochemistry*, 1978, **17**, 5460–5468.
36. Richards, F. M. and Vithyathil, P. J., *J. Biol. Chem.*, 1959, **234**, 1459–1464.
37. Wlodawer, A., Svensson, L. A. and Gilliland, G. L., *Biochemistry*, 1988, **27**, 2705–2717.
38. Kim, E. E., Varadarajan, R., Wyckoff, H. W. and Richards, F. M., *Biochemistry*, 1992, **31**, 12304–12314.
39. Connelly, P. R., Varadarajan, R., Sturtevant, J. M. and Richards, F. M., *Biochemistry*, 1990, **29**, 6108–6114.
40. Varadarajan, R., Connelly, P. R., Sturtevant, J. M. and Richards, F. M., *Biochemistry*, 1992, **31**, 1421–1426.
41. Varadarajan, R. and Richards, F. M., *Biochemistry*, 1992, **31**, 12315–12327.
42. Thomson, J., Ratnaparkhi, G. S., Varadarajan, R., Sturtevant, J. M. and Richards, F. M., *Biochemistry*, 1994, **33**, 8587–8593.
43. Burling, F. T., Weis, W. I., Flaherty, K. M. and Brunger, A. T., *Science*, 1996, **271**, 72–76.

ACKNOWLEDGEMENTS. We thank the Area Detector Facility located at the Molecular Biophysics Unit, Interactive Graphics Facility and the Supercomputer Education and Research Center at the Indian Institute of Science for use of their facilities. We thank Prof. M. R. N. Murthy, MBU for helpful discussions. Financial support for this work was provided by CSIR (E.M.R. No. 37(0913)/96/EMR-II) and DST (No. SP/SO/D-21/93). G. S. R. is a K. S. Krishnan fellow supported by the Department of Atomic Energy.