Biochemistry (Moscow), Vol. 67, No. 5, 2002, pp. 520-524. Translated from Biokhimiya, Vol. 67, No. 5, 2002, pp. 624-630.

REVIEW

Models for Protein Folding and Nature's Choice of Protein as Catalyst

R. H. Khan* and F. Khan

Interdisciplinary Biotechnology Unit, Aligarh Muslim University, Aligarh 202002, India; fax: 91-571-701081; E-mail: rizwanhkhan@hotmail.com

Received June 19, 2001
Revision received August 28, 2001

Abstract—The study of protein folding and unfolding pathways lends a fascinating dimension to protein biochemistry. Several models for protein folding have been postulated. Two powerful probes used in protein folding study are far UV-CD monitored stopped flow kinetics and pulse hydrogen exchange in conjunction with NMR. The formation of molten globule, which is an intermediate possessing secondary structure but not a well packed tertiary structure, is now emerging as a common feature on the folding pathway of many proteins. The molten globule is recognized by a class of molecules called chaperones which act as accelerators of protein folding. This article ends by elucidating why proteins are Nature's choice as catalysts.

Key words: protein folding, molten globule, biocatalysis

Protein folding is defined as the process in which neighboring and distant amino acids of a linear polypeptide come closer to each other and join to form a native structure with unique biological activity. Once the protein is synthesized only its folded form is functional, so the folding of the polypeptide is an important step in converting the genetic information into the functioning of the cell [1]. The first studies on folding of proteins were done in 1920 by Anson and Mirsky who concluded that denaturation of proteins is reversible [2]. In 1961 Anfinsen stated that the three dimensional structure of a protein is dictated by its primary sequence. The question arises why should one carry out protein folding studies. There are several reasons for this. First, expression of recombinant genes leads to the production of inclusion bodies which are insoluble and inactive forms. These proteins have to be made active, for which protein folding studies are required. Second, one of the major areas of studies these days is designing proteins with novel biological function, and for this purpose one should know how each amino acid will fold. Modification of proteins can also be carried out for which one should know the effect of altering the primary sequence and how this will affect the structure and folding. Third, protein folding studies help to understand diseases like Alzheimer's disease, scurvy, scrapie, and cystic fibrosis which are caused by mutation in protein causing the aggregation or misfolding of proteins. It also helps us to understand how some human proteins undergo conformational changes that render them pathogenic. The amyloid and prion diseases appear to result from conversion of one of the soluble and functional proteins into β-sheet rich quaternary structures that are often fibrillar, and this occurs at low pH or denaturing conditions [3]. This proves that the aqueous environment (pH, temperature, ionic strength, presence of chaotropic agent) also strongly influence the conformation adopted by a polypeptide. The cellular prion protein (PrPc) C-terminal structure is environmentally dependent and the N-terminal portion of PrPc is largely unfolded. Fourth, to predict the structure of proteins, techniques like X-ray crystallography are not enough, folding studies are also required. Fifth, structure function relationship of proteins and protein–protein interaction can be studied at the molecular level only.

REASON FOR ELUCIDATION OF PROTEIN FOLDING PATHWAY

The discrepancy between the calculated and experimental results obtained for protein folding is called the Levinthal paradox. Consider a small protein with 100 residues. Cyrus Levinthal calculated that if each residue can assume three different positions, the total number of structures is 3^{100} which is equal to 5×10^{47}. If it takes 10^{-13} sec to convert one structure into another the total search time would be $5 \times 10^{47} \times 10^{-13}$ sec which is equal to 1.6×10^{27} years. But real proteins fold within seconds or

* To whom correspondence should be addressed.

0006-2979/02/6705-0520$27.00 ©2002 MAIK "Nauka/Interperiodica"

minutes. This enormous discrepancy is called the Levinthal paradox [4]. Hence it was inferred that proteins cannot fold by a random search of all possible conformations, rather protein folding occurs by progressive stabilization of intermediates. The essence of protein folding is the retention of partially correct intermediates.

Partially folded intermediates can be detected, trapped, and characterized by rapid kinetic studies, trapping of disulfide bonded intermediates, pulse labeled NMR, site specific mutagenesis, and protein design.

MODELS OF PROTEIN FOLDING

Many models for folding have been proposed. They are based on theoretical considerations or phenomenological constructs and on computer simulation. Molecular dynamics (MD) simulations give a detailed description of the behavior of proteins during folding and unfolding processes [5]. However, these simulations give a more correct picture of the unfolding because to accelerate the events in simulation the temperature has to be raised, which corresponds to the unfolding conditions. MD operate with nearly folded protein only and point out that the most essential step in folding is formation of a relatively small critical folding nucleus in the unfolded phase. The interpretation of MD studies is not always reliable because a folding process must proceed via the same pathway as the reverse unfolding process only when both are held at the same conditions, but processes under different conditions are not obliged to follow the same route.

Framework model. The framework model was proposed by Ptitsyn [6]. According to this model, protein folding is hierarchial, that is during folding secondary structural elements are formed initially, which then constitute the framework for further formation of tertiary structure. Fluctuating segments of secondary structure first form and then coalesce so that the polypeptide chain becomes compact and the final tertiary structure forms subsequently.

It is said that a protein achieves its native structure avoiding a large number of alternatives because it folds step by step in such a way that the results of each step are not reconsidered at subsequent steps, but just fastened by them. The main point of this scheme is that three levels of protein structure—secondary structure, folding pattern, tertiary structure—are assumed to be formed as three subsequent stages of protein folding [7]. The first (burst) intermediates predicted by this scheme were revealed independently by two groups [8, 9]. The result showed that this intermediate is formed within a few milliseconds. It has usually 50-100% of native far UV-CD and therefore must have a very substantial amount of secondary structure [10]. Since it binds the hydrophobic probe (8-anilino-1-naphthalene sulfonate; ANS) it suggests that it has solvent accessible clusters of nonpolar groups. The second intermediate revealed usually forms within a second and is almost as compact as the native state [11]. The pulse hydrogen labeling technique elaborated by Baldwin and Roder has shown that this state has many stable α-helices and β-strands. Stable secondary structure forms after fluctuating secondary structure [12, 13].

Hydrophobic collapse model. In this model the hydrophobic interaction which can be described as water driving two nonpolar molecules together so that their contact with water is reduced, is thought to be the dominant interaction in determining the folding of protein. The nonspecific nature of hydrophobic interactions suggested that they could form more quickly than other more specific interactions. The first direct evidence for the hydrophobic collapse model has been found in the case of barstar [14].

Much study has been done on the protein–protein interaction and folding process in barstar (89 amino acid protein which is an intracellular inhibitor of bacterial RNAse or barnase) [15]. Unfolding kinetics induced by urea have been carried out in four different conditions, i.e., at 10°C, 25°C (in the presence and absence of 0.8 M KCl), and 40°C. The presence of 0.8 M KCl significantly stabilizes the native state with respect to unfolded state, suggesting that its effect is through strengthening of hydrophobic interaction or weakening of unfavorable electrostatic interaction. The hydrophobic interactions become weaker with decrease in temperature from 25 to 10°C. The burst phase loss in secondary structure at higher temperature but not at lower temperature is because hydrogen bonding interactions become weaker at 40°C. The formation of two burst phase unfolding intermediates with different far UV-CD and fluorescence properties have been suggested. One is devoid of substantial secondary structure resulting in reduced far UV-CD signal and the other has perturbed hydrophobic core resulting in reduced fluorescence signal. These two burst phase intermediates must be on two different unfolding pathways, each with different transition states.

Nucleation–condensation growth model. For a protein to achieve its native conformation in a biologically feasible time it has to sample a limited number of possible structures; this is achieved by nucleation, which reduces the number of searches for native protein structure [16, 17]. In the nucleation condensation growth model (NCGM), a region of the polypeptide chain serves as a nucleus for chain propagation to obtain the native state. The nucleation site is suggested to be around 8-18 amino acid residues long. The nucleus is expected to be unstable by itself and is therefore present only for a short while in the denatured state.

Jigsaw puzzle model. The assembly of structural elements can occur in many different ways analogous to the assembly of the pieces of the jigsaw puzzle [18]. There is no definite order to their assembly but this theory has

been discarded because it has been established that folding occurs along a few well defined sequential pathways with intermediates rather than a number of equally accessible pathways.

STUDY OF PROTEIN FOLDING

The earliest steps in protein folding are usually studied using rapid kinetic measurements. The earliest events of structure formation take place within the first 2-4 msec before they can be observed experimentally. Two powerful probes used to study early secondary structure formation are far UV-CD monitored stopped flow kinetics and pulsed hydrogen exchange in conjunction with NMR. Although CD provides a useful method for following secondary structure formation it does so without defining the origin of specific features. In contrast, pulsed hydrogen-exchange methods can detect the formation of specific secondary structure during folding [19]. These may not however detect the initial formation of isolated elements of secondary structure. The structure formed in the initial stages of folding appeared to be too labile to afford significantly intense labeling pulse used in the experiments.

It is difficult to tell whether the secondary structure is formed as a result of hydrophobic collapse or formation of secondary structure takes place before a hydrophobic collapse because methods for determination of a collapse by direct measurements are not readily coupled to rapid data acquisition systems [20]. In this context fluorescence spectroscopy study is very useful. The fluorescence intensity of tryptophan is used typically to monitor folding in most kinetic studies because it is sensitive to environment and its variation can be easily monitored [21]. It was seen that burst phase fluorescence changes within the dead time (5 msec) of measurement indicating that at least some side chains interactions involving tryptophan have already formed on this time scale.

Direct measurements of protein compactness by small angle X-ray scattering have been performed recently in both equilibrium [22] and kinetic experiments [23]. Measurements of the scattering intensity spectrum indicates that a protein is indistinguishable in terms of its compactness from the native state much before folding is complete. To compare the various kinetic and equilibrium intermediates it is better to use solvent than temperature induced denaturation as heating usually does not completely unfold proteins and leads to their aggregation while urea and guanidine hydrochloride (GuHCl) at room temperature unfold proteins more or less completely [24]. Urea induced denaturation of lysozyme and myoglobin was studied by column chromatography and it was seen that at small denaturant concentration only one peak of elution volume is observed which corresponds to the compact native molecule. As the denaturant concentration is increased a second peak is seen which is of the denatured unfolded molecule. This peak grows till the first one completely disappears. This phenomenon proves the all or none character of solvent induced denaturation of proteins. The all or none character of transition means that each molecule undergoes the transition as whole but it does not mean that there is the absence of any other transition. Two transitions were seen in carbonic anhydrase and β-lactamase induced by GuHCl. The first transition is the protein denaturation monitored by loss of its activity and rigid tertiary structure. The second transition can be monitored by change in elution volume and CD in the peptide region. This implies that there is a denatured state with compactness and secondary structure content close to those of native state but without activity and rigid tertiary structure. These properties are similar to that of a molten globule. A molten globule is an intermediate which possesses secondary structure but not a well packed tertiary structure. The formation of a molten globule is emerging as a common feature on the folding pathway of many proteins. It was first extensively characterized as an equilibrium intermediate [25, 26].

Low temperature denaturation gives a different picture. Here it is possible to observe three transition curves. The first is of protein denaturation, the second is the all or none transition between two denatured states, and the third shows that even after all the molecules have been transformed into less compact state, they continue to expand. The existence of three transitions shows the presence of at least two intermediates between the native and completely unfolded molecule. The more compact intermediate is almost as compact as the native protein and has high secondary structure and also binds ANS, that is it fulfills all the criteria of a molten globule while the compact intermediate is the partly folded state. It also has some secondary character but binds ANS weakly. This state also is similar to the first kinetic intermediate. Hence it appears that there is similarity between the properties of equilibrium and kinetic molten globule states.

Spectroscopic techniques measure the structural changes only in the millisecond time scale so the folding occurring earlier than these cannot be detected. This problem can be circumvented by changing the physical or chemical conditions of the reactions like lowering the temperature [27, 28]. The rate of folding does not depend specifically on its size or structure and there is also no connection between stability and folding. The transition state in protein folding is compact and is more like the native than the unfolded form [29]. This state is at the highest energy level on the pathway of folding and follows the classical transition theory of chemical reactions even though it is composed of mainly weak interactions as opposed to chemical reactions involving the making or breaking of covalent bonds. Even if multiple transition states are present they are equivalent in energy. It was also seen that the transition state for the folding and unfolding pathways are not same [30]. It seems that protein folding

follows kinetically accessible pathways which are much less than the total number of possible pathways because the local free energy minimum is at higher energy level than the global free energy minima; otherwise, if a native protein structure was of the lowest free energy its formation would take a very long time.

The folding kinetics has two distinct phases. The fast phase in which all the main structural changes like the hydrophobic collapse, hydrogen bond formation, and packing of secondary structure take place. The slow phase is seen in certain proteins involving isomerization about proline imide bonds [31, 32]. The *trans* form is thermodynamically favored in the unfolded form and the slow phase of protein folding is because of the slow isomerization of non-native *trans* form of proline to native *cis* form during folding.

It has been very difficult to study the folding pathways because most of the intermediates at equilibrium are highly unstable. Site-directed mutagenesis is a potential tool for breaking down the cooperativity of folding process so that the intermediates can be accumulated at equilibrium and made easier to study [33]. The changes in cooperatively along the folding pathway can be measured by double mutant cycles [34, 35]. This cycle estimates the extent of interaction between two residues in a folded protein. In this method two interacting residues are identified and mutated both independently and simultaneously. Then the difference in the free energy of unfolding ($\Delta\Delta G$) of the mutant proteins are measured with respect to the wild type. The comparison of the difference of $\Delta\Delta G$s of two single mutant with $\Delta\Delta G$ of double mutant indicated the extent of interaction between the two residues under study. Protein engineering methods have been used to characterize the transient intermediate on the folding pathway [36]. Folding pathways of certain proteins like bovine pancreatic trypsin inhibitor (BPTI) which require the formation of specific disulfide bonds for its folding have been studied and trapped intermediates have been characterized by selectively mutating the Cys residues and measuring the rate of disulfide bond formation [37, 38].

ACCELERATORS OF PROTEIN FOLDING

The molten globule is recognized by a class of molecules called "chaperones" [39] which are involved in the insertion of proteins into the membrane. There are also suggestions that some proteins like insulin may function in the molten globule state [40].

When proteins are expressed as insoluble aggregates (inclusion bodies) the recovery of functional proteins is required. There are three set of conditions necessary for successful recovery of functional proteins after denaturation. One, the folded protein should be most stable. Cosolvents like glycerol, ethylene glycol, and sucrose, anions like phosphate and sulfate, and cations like MES (2-[N-morpholino]ethanesulfonic acid) can stabilize protein markedly against denaturation. These do not affect folding rate constant but decrease the unfolding rate constant hence stabilizing the protein. Two, the kinetic barriers that slow down or block the folding pathway should be minimized. Three, the intermolecular aggregation between folding intermediates should be reduced to acceptable limits. This is done by a group of protein chaperones, proteins GroEL and GroES have been shown to dramatically increase the yield of refolding in proteins [41]. Chaperones along with PDI (protein disulfide isomerase) and PPI (peptidyl propyl *cis-trans* isomerase) are some factors that increase the yield and the rate of folding. The rate of folding will be inverse to the concentration of denaturant. For certain proteins whose folding rate is limited by proline *cis-trans* isomerization, the rate is increased by adding PDI [42]. A major factor in stabilizing aggregates appears to be hydrophobic interactions; therefore, factors that can decrease the strength of such interactions or increase the solubility of other parts of a protein chain will abolish problems of low yield stemming from aggregation.

PROTEINS: NATURE'S CHOICE AS CATALYST

Lastly, it is interesting to note that evolution preferred proteins as cellular catalysts over any other macromolecule. The reason cannot be attributed to composition alone but the reason lies in the structure of these molecules. Protein folding is a highly cooperative process as compared to folding in RNAs which are the other catalytic molecules of nature [43]. RNA secondary structure is extremely stable in the absence of tertiary structure in contrast to proteins. Also, side chains are on the inside of RNA double helix while amino acid side chains face outwards in a protein secondary structure allowing cooperative hydrophobic interactions to form tertiary structures, which is a major governing force in protein folding. High cooperativity and weak non-complementary interactions in protein folding may significantly contribute to the fact that molecular evolution selected protein as catalyst.

Finally, we would like to suggest possible factors for selecting the proteins as catalyst by Nature, which are as follows. First, the bulk of the catalyst (proteins) structure is rigid and the buried hydrophobic amino acids provide the favorable environment for the interaction with their substrate. Second, only a few charged amino acids are responsible for catalysts with a broader range of pH optima. Since only amino acid can cover the entire pH range for catalysis, i.e., for acidic aspartic acid/glutamic acid and for basic histidine/lysine, it may be the possible factor for selecting protein over some other macromolecule as a catalyst.

We are grateful to Dr. M. K. Ali for cooperation during the preparation of this manuscript. Financial support by CSIR grant No. 38 (978) 99EMR II is gratefully acknowledged. The authors are also highly thankful for the facilities provided at A. M. U., Aligarh.

REFERENCES

1. Nath, U., and Udgaonkar, J. B. (1997) *Curr. Sci.*, **72**, 180-191.
2. Anson, M. L., and Mirsky, A. E. (1939) *J. Phys. Chem.*, **35**, 185.
3. Kelly, J. W. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 930-932.
4. Levinthal, C. (1968) *J. Chem. Phys.*, **65**, 44-45.
5. Finkelstein, A. V. (1997) *Prot. Eng.*, **10**, 843-845.
6. Ptitsyn, O. B. (1994) *Prot. Eng.*, **7**, 593-596.
7. Ptitsyn, O. B. (1987) *J. Protein Chem.*, **6**, 273-293.
8. Gilmanshin, R. I., and Ptitsyn, O. B. (1987) *FEBS Lett.*, **233**, 327-329.
9. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., and Nagamura, T. (1987) *FEBS Lett.*, **221**, 115-118.
10. Segawa, S., and Sugihara, M. (1984) *Biopolymers*, **23**, 2473-2488.
11. Dolgikh, D. A., Kolonuen, A. P., and Ptitsyn, O. B. (1984) *FEBS Lett.*, **165**, 88-92.
12. Baldwin, R. L., and Roder, H. (1991) *Curr. Opin. Biol.*, **1**, 218-220.
13. Roder, H., Elove, G. A., and Englander, S. W. (1988) *Nature*, **335**, 700-704.
14. Agashe, V. R., Shastry, M. C. R., and Udgaonkar, J. B. (1995) *Nature*, **377**, 745-755.
15. Zaidi, F. N., Nath, U., and Udgaonkar, J. B. (1997) *Nat. Struct. Biol.*, **4**, 1016-1024.
16. Matheson, R. R., and Scheraga, H. A. (1978) *Macromolecules*, **11**, 819-835.
17. Tanford, C. (1968) *Adv. Prot. Chem.*, **23**, 121-282.
18. Harrison, S. C., and Durbin, R. (1973) *Proc. Natl. Acad. Sci. USA*, **82**, 4028-4030.
19. Wetlaufer, D. B. (1985) *Proc. Natl. Acad. Sci. USA*, **70**, 697-701.
20. Dill, K. A., Fiebig, K. M., and Chan, H. S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 1942-1946.
21. Eftink, M. R. (1994) *Biophys. J.*, **66**, 482-501.
22. Kataoka, M., Nishii, I., Fujisawa, T., Tokugawa, F., and Goto, Y. (1995) *J. Mol. Biol.*, **249**, 215-228.
23. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O., and Tsuruta, H. (1995) *Science*, **270**, 480-488.
24. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1992) *Biochemistry*, **29**, 9343-9352.
25. Fink, A. L. (1985) *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 492-522.
26. Ptitsyn, O. B., Pain, R. M., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990) *FEBS Lett.*, **262**, 20-24.
27. Jones, C. M., Henry, E. R., Hu, Y., Chen, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W. A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11860-11864.
28. Pascher, T., Chesick, J. P., Winkler, J. R., and Gray, H. B. (1996) *Science*, **271**, 1558-1560.
29. Salahuddin, A. (1980) *J. Sci. Ind. Res.*, **39**, 745-751.
30. Houry, W. A., Rothwarf, D. M., and Cheraga, H. A. (1995) *Nat. Struct. Biol.*, **2**, 495-503.
31. Kiefhaber, T., Grunert, H. P., Hahn, U., and Schimd, F. X. (1990) *Biochemistry*, **29**, 6475-6480.
32. Nakano, T., Antonino, L. C., Fox, R. O., and Fink, A. L. (1993) *Biochemistry*, **32**, 2534-2541.
33. Woodward, C. K. (1994) *Curr. Opin. Struct.*, **4**, 112-116.
34. Carter, P. J., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1984) *Cell*, **38**, 835-840.
35. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Golmanshin, R. I. (1991) *Biopolymers*, **31**, 119-128.
36. Matouschek, A., Serrano, L., and Fersht, A. R. (1992) *J. Mol. Biol.*, **224**, 819-835.
37. Creighton, T. E. (1992) *Nature*, **356**, 194-195.
38. Darby, N. J., and Creighton, T. E. (1993) *J. Mol. Biol.*, **232**, 873-896.
39. Martin, J., Langer, T., Boteva, R., Scharamell, A., Horwich, A. L., and Hartl, F.-U. (1991) *Nature*, **352**, 36-42.
40. Hua, Q. X., Ladbury, T. E., and Weiss, M. A. (1993) *Biochemistry*, **32**, 1433-1442.
41. Goloubinoff, P., Christeller, J. T., Gatenby, A. A., and Lorimer, G. H. (1989) *Nature*, **342**, 884-889.
42. Nall, B. T. (1994) in *Mechanisms of Protein Folding* (Pain, P. H., ed.) Oxford University Press, N. Y., pp. 80-103.
43. Csermely, P. (1997) *Trends Biochem. Sci.*, **22**, 147-149.