Protein folding revisited

M. Yahiya Khan

'Protein chemistry is now dead'. This was a comment made in the early eighties by one of the members of the UGC team visiting our department. This by no way means that the honourable member of the UGC team was having any bias towards protein chemistry or he was not personally happy with the members of the faculty working in that area. Instead, he was simply reflecting the general opinion of the large number of 'scientists' about the subject at that time. There were a variety of reasons for that widely held view. The stormy emergence of biotechnology, the initial success in cloning foreign genes into useful hosts and the enthusiasm for studies on gene structure and expression were some of the factors that over-shadowed protein chemistry for several years.

The honeymoon with gene cloning and expression has however, proven to be short-lived. Despite numerous successes, difficulties soon emerged in recovering correctly folded, active proteins from cloned genes[1]. This pursued a group of researchers led by Jonathan King at MIT, USA, to organize the first ever three day conference on the problem of protein folding[2]. Unexpectedly, the meeting was a great success, heralding a new era in the history of protein chemistry in general and protein folding in particular. Since then, a special seminar on protein folding has become a regular feature of the annual meetings of the American Association for the Advancement of Science. The latest seminar of the series was held at New Orleans in February this year. Protein folding soon became a fashion of scientific journalism, so much so, in addition to scientific journals, *The New York Times* published a popular article on the subject[3].

Genetic engineering is as much responsible for the new fashion in protein folding as it was behind its eclipse almost two decades ago. The approach towards the study of protein folding is changing dramatically. This brief account is restricted to recent advances in the field. For earlier studies on the subject, the readers are referred to other recent reviews[4-7].

The problem of protein folding

The biological activity of a protein depends on its three-dimensional structure which is derived by the folding of the linear amino acid polymer synthesized on ribosomes. Protein folding is generally a fast reversible reaction. Folded proteins have only a marginal stability over their unfolded counterparts. The physical forces that determine the fate of unfolding–refolding reactions include, but are not limited to, the hydrophobic interaction, the van der Waals interaction, electrostatic interactions, hydrogen bonds and covalent cross links. Since the magnitude of these forces is influenced by environmental conditions, protein stability depends on factors such as temperature, pressure, pH, extraneous agents (like specific ligands) and ionic strength. In short, the folded conformation of proteins is the result of the simultaneous interaction of the constituent amino-acid residues of the protein polypeptide chain with each other as well as with the environment in which they are present. The biggest challenge of a modern protein chemist is therefore, to determine the contributions of individual amino acids to the stability of a protein as a function of environmental conditions.

The improvement of genetic methods, have greatly enhanced the reliability of the measurements of the stabilizing contribution of specific interactions in proteins. Studies involving mutational substitutions of critical amino acids in several proteins have revealed that the proteins are very tolerant to many of these substitutions[4]. For instance, more than 50% of 323 substitutions at 90 sites in lac repressor of *E. coli* yielded a wild type phenotype[8].

The most important and perhaps controversial question one can ask at this stage would therefore be that, should we continue to believe that primary structure alone determines the tertiary structure of a protein? The answer is both, yes and no. Yes, because we still do not have enough evidences to prove it otherwise. No, because, as mentioned above, (i) some problems in recovering folded active proteins from cloned genes have emerged, (ii) excessive substitutions in the amino-acid sequence have not led to any change in the properties of certain proteins, and (iii) helper protein molecules or chaperones have been discovered which are suggested to be involved in directing the correct folding of certain proteins. The following discussion would be centred around the formation of inclusion bodies—the denatured proteins generated within the optimally native conditions within the cell and the probable role of molecular chaperones in protein folding.

Inclusion bodies

The term 'inclusion bodies' generally refers to non-functional protein aggregates generated within cells under physiological conditions[1]. Electron microscopic studies have shown that inclusion bodies are amorphous, dense structures without a cell membrane. They vary in size but can be morphologically identified in the cytoplasm.

There are variety of reasons for the formation of inclusion bodies (for review see ref. 1). The change in the structure of a protein due to covalent damage or modification, for instance, may lead to the formation of inclusion bodies. Similarly, mutant proteins may also have a greater tendency to form the inclusion bodies. The most important factor which appears to be of direct relevance to a modern protein engineer is, however, the environment where protein folding occurs. An alteration of the expression conditions of the wild type *E. coli* proteins within the host cell for instance, has resulted in the formation of inclusion bodies[9,10]. Formation of inclusion bodies from products of an animal gene cloned in a bacterium can thus be attributed to altered expression conditions.

There are two possible ways by which an alteration in the expression conditions might lead to the formation of inclusion body aggregates. Firstly, the change in environment may result into the generation of an incorrectly trans-

cribed or translated sequence. Alternatively, the change in the intracellular conditions might effect the folding characteristics of a protein in such a way that the formation of inactive aggregrates is preferred over the formation of functional native protein. That the change in sequences in proteins-forming inclusion bodies is highly unlikely, if at all not completely ruled out, is shown by regeneration of active proteins from their respective inclusion body aggregates by *in vitro* unfolding–refolding experiments (for review see ref. 1). The problem of inclusion body formation thus appears to be associated with the conformation and not with the sequence of the protein. There are several examples which support this contention. London *et al.*[11] for instance, found that the renaturation conditions were critical parameters in determining the fraction of the active native proteins recovered as the end products. They went on to propose that the precursors in the process of inclusion body formation were partially folded intermediates formed during the refolding at a particular denaturant concentration. That the generation of the active, native state of a protein depends on the nature of the partially folded intermediates in the folding pathway is demonstrated by our recent studies on urea inactivation of goat spleen cathepsin B[12]. After treatment with different urea concentrations (presumably giving different intermediate conformational states) and subsequent refolding by dilution, several cathepsin B conformers with varying activity were obtained[12].

From the forgoing remarks, it can be concluded with reasonable amount of confidence that inclusion body formation is the result of intracellular protein folding which is somewhat different from the 'self-assembly' leading to the formation of active native protein. I describe next the factor(s) that might be responsible for the preference of the *in vivo*-protein-folding-intermediates to form active folded protein over the inactive inclusion bodies.

Molecular chaperones

Although there are several factors (e.g. ions, cofactors, intracellular environment, etc.) which might play crucial roles in leading to correct protein folding *in vivo*, molecular chaperones are thought to be the most essential protein factors that help in biogenesis of macromolecular structures[1,13,14]. They appear to prevent the formation of certain intermediate structures that might lead to acquisition of nonfunctional protein conformation(s). The precise mechanism of chaperone action is not known till date. It is however, conceivable that chaperones might recognize newly synthesized (presumably unfolded) proteins and associate with them transiently until they are assembled into their correct native structures[13,14].

Our knowledge about molecular chaperones has primarily came from studies carried out during the last decade (for review see ref. 13). They can be broadly grouped into three classes namely, the nucleoplasmins, the chaperonins and the Bip–hsp 70 class. Nucleoplasmins are suggested to be involved in nucleosome assembly while the chaperonins, which may have prokaryotic, mitochondrial or plastid origins, are proposed to be active in DNA replication, protein transport, phage assembly and protein folding. Bip–hsp 70 class representing heat-shock proteins and their homologues, and immunoglobulin heavy chain binding proteins, are known to be crucial in processes like protein transport, and assembly within endoplasmic reticulum and mitochondria. The first crystal structure of a chaperone protein, Pap D which mediates assembly of pili in *E. coli* has been recently published[15]. Studies like this are bound to provide new insights about the mechanism(s) of chaperone action.

Perspective and conclusion

In conclusion we can safely say that protein chemistry is growing 'younger' everyday. More crucially new findings have not been able to even alter the old ideas, leave alone lead to their complete rejection. The technologies have, however, opened new frontiers for protein chemists where they can now play a leading role in providing the information necessary for solving the problems faced by protein engineers in recovering active proteins. The cause and mechanism of inclusion bodies formation, possible ways to avoid their formation, molecular chaperones and the mechanism of their action, and the role of the environment and other factors on protein folding are some of the areas where protein chemists will find themselves engaged in the years and decades to come.

1. Mitraki, A. and King, *J. Bio/Technology*, 1989, **7**, 690.
2. *The Protein Folding Problem*—A special 3-day seminar at the AAAS Annual Meeting, Boston, USA, February 12–14, 1988.
3. Gleick, J., *The New York Times*, June 14, 1988, p. C1.
4. Alber, T., *Annu. Rev. Biochem.*, 1989, **58**, 765.
5. Wright, P. E. Dyson, H. J. and Lerner, R. A., *Biochemistry*, 1988, **27**, 7667.
6. Schellman, J. A., *Annu. Rev. Biophys. Biophys. Chem.*, 1987, **16**, 115.
7. Dill, K. A., *Protein Engineering* (ed. Oxender, D. L.), C. F. Fox, New York Liss, 1987, pp. 187 192.
8. Miller, J. H. *et al.*, *J. Mol. Biol.*, 1979, **131**, 191.
9. Gribskov, M. and Burgess, R. R., *Gene*, 1983, **26**, 109.
10. Botterman, J. and Zabeau, M., *Gene*, 1985, **37**, 229.
11. London, J., Skrzynia, C. and Goldberg, M., *Eur. J. Biochem.*, 1974, **47**, 409.
12. Agarwal, S. K. and Khan, M. Y., *Biochem. J.*, 1988, **256**, 609.
13. Ellis, R. J. and Hemmingsen, S. M., *Trends Biochem. Sci.*, 1989, **14**, 339.
14. Sambrook, J. and Gething, M. J., *Nature*, 1989, **342**, 224.
15. Holmgren, A. and Branden, C. I., *Nature*, 1989, **342**, 248.

M. Yahiya Khan is in the Department of Biochemistry, School of Life Sciences, North-Eastern Hill University, Shillong 793 014, India