



Protein Folding or Misfolding: A Problem and its Consequences

Agarwal SK* and Singh S

Department of Biochemistry, University of Lucknow, India

*Corresponding author: Sudhir K Agarwal, Department of Biochemistry, University of Lucknow, India, Email: agarwalsk5@rediffmail.com

Short Communication

Volume 5 Issue 1

Received Date: June 12, 2020

Published Date: June 26, 2020

DOI: 10.23880/ijbp-16000176

Abstract

Protein folding is a single most important process in biology whose complexity within living cell has been appreciated. Any deviation from the predetermined folding pathways might lead to the formation of 'non-native' conformations. Furthermore, mutations cause misfolded, non-functional form of proteins to accumulate. In spite of the tremendous progress made in the area of protein engineering, we can not reliably predict structural motifs in proteins from their amino acid sequence alone. Thus the real challenge is to work out the guidelines that might enable to predict how a protein would fold.

Keywords: Protein folding; Protein misfolding; Factors; Nucleation; Consequences

Introduction

'Amino acid sequence of a protein determines its unique three dimensional structure' – this was demonstrated in 1973 by Anfinsen [1] and has since been confirmed many times. For example, formation of a cleft (active site) in an enzyme or of a particular three-dimensional (3D) structure in a hormone or receptor is possible only when the protein acquires its folded structure. However, we do not know how a given portion of a polypeptide undergoes α -helical folding while another part acquires β -structure and yet some portion remains unstructured. Subsequently, how these ordered conformations interact with each other to form a 3D structure. Thus, how do enzymes or proteins, unaided by any other cellular component, acquire their functional structure is a vital issue whose resolution is the first necessary step in gaining a meaningful understanding at a molecular level.

The purpose of this article is not to review all the experimental results and theoretical treatments which have already been described earlier [1-7], but to focus attention primarily on relatively recent developments which can provide information on the possible mechanism of protein folding and on the physical constraints that are likely to govern (i) the stability of the native protein structure, (ii) in directing the correct folding of certain proteins and (iii) inherited disorders due to defective dynamics of protein

folding.

The Problem of Protein Folding

The folding of a polypeptide chain from a linear random state to an ordered structure is the final step of protein synthesis. Each protein acquires its 3D structure in an autonomous way according to specific sequential order of amino acids in the protein chain. The mechanism, by which the final structure acquires, is still an enigma. Since specific sequence of amino acids alone can satisfy the requirements for correct folding to attain its functional entity [1] and if so, the alternation in a sequence should change the character (structure and function) of a protein but excessive substitution in the amino acid sequence showed that there is no change in the properties of certain proteins [2]. The problem is that the rules governing the folding process i.e. the relationship between some groups of amino acids and certain final structural arrangements are not yet known. Therefore, guessing is not the way to predict a structure for a given sequence of amino acids. Although three dimensional structures of a number of proteins have been determined by X-ray methods, a statistical analysis of these structures has not been successful in developing predictive rules to relate amino acid sequence to 3D structures [3]. Thus the 'protein-folding problem' is one of the central mysteries in protein chemistry for the last 5-6 decades [3,4].

Theories of Protein Folding

The existing theories of protein folding mechanism have been developed over the past several years. These are 'thermodynamic control' and 'kinetic drive'. After the discovery by Anfinsen, it is commonly accepted that the tertiary structure of a protein is determined by the relative free energies of all its possible conformations [1], and if it is true then the folding process must be started from one of the large number of possible conformations of the protein molecule. This implies further that there are enormous numbers of folding pathways, yet the end product i.e. the native state (lowest free energy) is the same. Thus, there is no path restriction in the thermodynamic control of protein folding. However, this has been ruled out due to the fact that the random search for the thermodynamically stable conformations is a process too slow to account for refolding of proteins in a few seconds [5]. Therefore, the arguments goes in favor of kinetic drive of protein folding which says that only a small fraction of all the possible conformational states (intermediates) are analyzed on the way to the native structure. In other words, the protein folding is path dependent and it is governed by kinetic considerations [6-9]. Overall, the existence of intermediates in the folding pathways of a polypeptide is widely accepted.

Folding of Polypeptide Chains and their Constraints

Although the molecular information to attain the final biologically active native structure is contained in the primary covalent structure of the protein [1,2], the driving force to attain the desired structure mainly arise from intramolecular non-covalent segment-segment interaction and also the interaction of the constituent amino acid residues with the surrounding solvent medium [10]. In fact a protein molecule becomes stable when it exists under conditions similar to those for which it was selected, the so called native state or physiological state. To work out this crucial phenomenon, a number of workers studied the unfolding-refolding reaction of several proteins including enzymes *in vitro* [8,9,11] as it is not possible to study the folding process during the synthesis of a protein in the cell. From the hierarchy of protein structure one might expect the folding of nascent peptide chain occurs as a co-translational events involving short range, nearest neighbor (formation of secondary structure as first step) and long range interactions (formation of super secondary and tertiary structures as a series of subsequent steps). These interactions are possible because of the unique structural organization of each amino acid. However, the contribution of individual amino acids towards the stability of a protein as a function of environmental conditions is not yet known and thus may be a challenging task for a modern protein chemist.

Nucleation-Directed Protein Folding

A protein molecule acquires native structure within few minutes or even less, probably through 'nucleation' which dramatically reduces the number of searches for native protein structures [12-14]. Wetlaufer and his group provided experimental evidences after observing rapid regeneration of enzymatic activity from inactive reduced lysozyme in a mixture of reduced and oxidized glutathione [13,15]. The rate limiting step in the regeneration process was neither thiol oxidation nor disulphide reshuffling, but the folding of the protein itself. Moreover, the existence of independently folded distinct structural regions (domains) in several globular proteins having a single polypeptide chain and the continuity of protein chain within them suggest that the nucleation may occur independently in separate parts of protein molecules [13]. This has confirmed by the study that the NH₂-terminal segment of chicken ovomucoid has a far higher rate of refolding than the fragments obtained from other regions of the polypeptide chain [9]. A nucleation site (8-18 amino acid residues) represents short segment(s) of a protein chain having a tendency to form a 'native format' and its formation greatly hastens the self-assembly of the protein. The nucleation process may be accomplished by the formation of a specific pocket of non-polar amino acid residues which is stabilized by hydrophobic interactions. Besides the involvement of all range interactions, hydrogen bonding also begins to contribute significantly to the stability of the nucleation site [12-15].

Factors Involved in Protein Folding

Thermodynamic stability of proteins depends on environmental conditions and even under the most favorable condition the folded native state is stabilized only by about 5-15 kcal/mole [16,17]. Thus the environment (solvent medium) plays a major role and believed to be the driving force in the final step of protein folding because the intramolecular non-covalent interactions themselves are modulated by the properties and composition of solvent medium. Hence, the parameters like pH, ionic strength, temperature and presence of specific ligands such as metal ions, cofactors or prosthetic groups are important factors on which the stability of a protein depends. For example, ovomucoid, an egg white protein, showed single step transition (N \leftrightarrow D) on heat-induced unfolding at pH 7.0 but it was resolved in two-step transition (N \leftrightarrow X \leftrightarrow D) at pH 5.0 [9]. Similarly, the presence of heme is known to facilitate the refolding of hemoglobin chains to their native conformation [18]. Likewise, a cofactor might be required exclusively for the folding and assembly of an enzyme without being necessary for the function of the native protein. It is therefore apparent that the stable (folded) conformation of protein is the result of the simultaneous interactions 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.

Role of Inclusion Bodies and Molecular Chaperones

The inclusion bodies generally represent the non-functional protein aggregates generated within cells under physiological conditions due to covalent damage or modifications, mutation and the change of environment [19,20]. The formation of inclusion bodies is the result of intracellular protein folding which is somewhat different from the self-assembly leading to the formation of active native proteins. The mechanism of inclusion body formation is not yet completely understood, it is, however, certain that 'non-native' interactions between the constituent amino acid residues and their environment are the primary cause of this vital phenomenon [19]. On the other side, molecular chaperones is one of the most essential protein factors that facilitate folding, assembly and targeting of nascent chain *in vivo* and prevent aggregation in refolding assays *in vitro* [20-22]. Although the precise mechanism of chaperone action is not yet known, they are believed to recognize the features present in incomplete folded polypeptide chains providing the time needed for folding interactions to drive the functional, native structure. Among three groups of molecular chaperones, nucleoplasmins involve in nucleosome assembly while chaperonins are active in DNA replication, protein transport, assembly and protein folding. GroEL, an *E. coli* chaperonin, binds to a peptide of rhodanese, a mitochondrial protein whose *in vitro* refolding is greatly enhanced by addition of GroEL with GroES and ATP [21,22]. Bip-hsp 70 class representing heat-shock proteins are known to be crucial in processes like protein transport and assembly within endoplasmic reticulum and mitochondria. The elucidation of 3D structure of a chaperone protein, namely Pap D from *E. coli*, has revealed that its polypeptide chain folds into two immunoglobulin type domains that are homologous in sequence to the human lymphocyte differentiation antigen Leu-1/CD 5 [17].

Defective Dynamics of Protein Folding

The protein misfolding causes varieties of diseases by 'loss-of-function' mutation (due to improper folding, degradation or localization) and disorders resulting from 'gain-of-function' mechanism (mutation that cause a toxic novel function, dominant-negative mutation and amyloid accumulation) [23].

Many cellular degradation systems like ERAD or autophagy have play important role in preventing the accumulation of non-functional misfolded proteins. For

example, cystic fibrosis initiated by the mutation, is mainly caused due to the deletion of a Phe residue at position 508 in cystic fibrosis transmembrane conductance regulator (CFTR), a plasma membrane chloride channel. This mutation results the protein to be misfolded and targeted for degradation but due to the disrupted function of the chaperone system, the mutant CFTR is escaped from the degradation mechanism. Upon knockdown of AHA1, a co-chaperone that, together with HSP 90, alters the maturation of CFTR, 'CFTR Phe508' is not only more stable, but partially functional [24]. Another related illustration of this category of protein-folding disease is "Gaucher's disease" [25] which are caused by a range of mutations in β -glucosidase, lysosomal enzyme having a role in metabolism of the lipid glucosylceramide [26]. For the second fact (totally devoted to the proper trafficking of proteins into subcellular localization), it is essential that the proteins must be properly folded. However due to mutations, the correct folding of proteins is destabilized and leads to improper localization. A dysfunction can arise via loss of function of the protein at its appropriate location and secondary by gain-of-function toxicity if it stores in an incorrect location [23]. A mutated α -1 antitrypsin, a secreted protease inhibitor, leads to emphysema in a recessive loss-of-function manner and liver damage in a dominant gain-of-function manner [27].

Another mode of protein misfolding is that result a disease through dominant-negative mechanism, which occurs when a mutant protein antagonizes the function of the wild-type (WT) protein, causing a loss of protein activity even in a heterozygote [23]. In case of epidermolysis bullosa simplex, an inherited connective tissue disorder, mutant forms of the keratin proteins KRT5 and KRT14 lead to severe blistering of the skin in response to injury [28]. Other canonical instance of dominant-negative mutations that involve protein misfolding and affect individuals to disease is the homotetrameric transcription factor p53 [29]. One of the most common genetic alterations seen in cancer is due to mutations in p53 [23]. An example of gain of toxic function is apolipoprotein E (APOE), a lipid transport molecule. At least one copy of the APOE4 allele is found in 65-80% of individuals with Alzheimer's disease [30]. Amyloidogenic (insoluble fibrous protein aggregates) proteins lead the problem in amyloid-related diseases which are classified on the basis of presence of similar toxic conformations. The formation of such protein conformations can lead to neurodegenerative disorders like AD, Parkinson's disease and Huntington's disease [31].

Conclusion

The magic of folding of a protein molecule from structureless random state to an ordered structure represents a complex process and several factors involved in

this purposes are still yet to be well resolved to work out the guidelines which can explore the fundamental questions of protein folding, structure and function, and synthesis of new proteins for medical and industrial applications. It can simply be defined that initiation of folding start from nucleation site, followed by long range interactions and then rearrangement for functional identity. The possibility of the involvement of pro-sequences in the protein folding process can't be ruled out. The role of the environment, molecular chaperones and their action, the cause and mechanism of inclusion bodies formation, possible ways to avoid their formation, the causes of improper folding (misfolding) leading to certain disorders and other factors on protein folding are some of the areas where protein chemists will find themselves engaged in the years.

Acknowledgements

This work was supported by Department of Higher Education, Government of Uttar Pradesh, for Centre of Excellence in Biochemistry. The infrastructural facilities provided by University of Lucknow, Lucknow, are also gratefully acknowledged.

References

1. Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181(4096): 223-230.
2. Albert T (1989) Mutational effects on protein stability. *Ann Rev Biochem* 58: 765-798.
3. Subramaniam E (1990) Protein crystallography and protein engineering. *Curr Sci* 59(15): 728-732.
4. Richards FM (1991) The protein folding problem. *Sci Ame* 264(1): 54-63.
5. Garel JR, Baldwin RL (1973) Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc Natl Acad Sci USA* 70(12): 3347-3357.
6. Baldwin RL (1975) Intermediates in protein folding reactions and the mechanism of protein folding. *Ann Rev Biochem* 44: 453-475.
7. Goldberg ME (1985) The second translation of the genetic message: protein folding and assembly. *Trends Biochem Sci* 10(10): 388-391.
8. Agarwal SK, Khan MY (1988) A probable mechanism of inactivation by urea of goat spleen cathepsin B. *Unfolding and refolding studies. Biochem J* 256(2): 609-613.
9. Das BK, Agarwal SK, Khan MY (1991) Unfolding-refolding behavior of chicken egg white ovomucoid and its correlation with the three domain structure of the protein. *Biochem Biophys Acta* 1076(3): 343-350.
10. Nemethy G, Peer WJ, Scheraga HA (1981) Effect of protein-solvent interactions on protein conformation. *Ann Rev Biophys Bioeng* 10(10): 459-497.
11. Agarwal SK, Salahuddin A (1987) Domain II + III of bovine serum albumin: Isolation and its characterization. *J Biosci* 12(3): 191-202.
12. Salahuddin A (1980) Self-assembly of native protein structure. *J Scient Ind Res* 39: 745-751.
13. Wetlaufer DB (1990) Nucleation in protein folding confusion of structure and process. *Trends in Biochem Sci* 15(11): 414-415.
14. Chaturvedi SK, Siddiqi MK, Rizwan PA, Khan H (2016) Protein misfolding and aggregation: Mechanism, factors and detection. *Process Biochem* 51(9): 1183-1192.
15. Johnson ER, Kyung Ja O, Wetlaufer DB (1976) Formation of three-dimensional structure in protein fragments. *J Biol Chem* 251(10): 3154-3157.
16. Pivalov PL, Gill SJ (1988) Stability of protein structure and hydrophobic interaction. *Adv Protein Chem* 39: 191-234.
17. Khan MY (1992) Protein folding: retrospective and prospective. *Indian J Biochem Biophys* 29(4): 311-314.
18. Leutzinger Y, Beychock S (1981) Kinetics and mechanism of heme-induced refolding of human alpha-globin. *Proc Natl Acad Sci USA* 78(2): 780-784.
19. Mitraki A, King J (1989) Protein folding intermediates and inclusion body formation. *Biotechnology* 7(7): 690-697.
20. Khan MY (1990) Protein folding revisited. *Curr Sci* 59(15): 723-724.
21. Thirumalai D, Lorimer GH (2001) Chaperonin mediated protein folding. *Ann Rev Biophys Biomol Struct* 30(1): 245-269.
22. Chakrabarti S, Hyeon C, Ye X, Lorimer GH, Thirumalai D (2017) Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium. *Proc Natl Acad Sci USA* 114(51): 10919-10927.
23. Valastyan JS, Lindquist S (2014) Mechanism of protein-folding diseases at a glance. *Dis Model Mech* 7(1):9-14.
24. Wang XD, Venable J, LaPointe P, Hutt DM, Koulov AV, et

- al. (2006) Hsp90 cochaperone Aha1 down regulation rescues misfolding of CFTR in cystic fibrosis. *Cell* 127(4): 807-815.
25. Cox TM, Cachon-Gonzalez MB (2012) The cellular pathology of lysosomal diseases. *J Pathol* 226(2): 241-254.
26. Grabowski GA (2008) Phenotype, diagnosis, and treatment of Gaucher's disease. *Lancet* 372(9645): 1263-1271.
27. Perchiacca JM, Ladiwala AR, Bhattacharya M, Tessier PM (2012) Structure-based design of conformation- and sequence-specific antibodies against amyloid β . *Proc Natl Acad Sci USA* 109(1): 84-89.
28. Chamcheu JC, Siddiqui IA, Syed DN, Adhami VM, Liovoic M, et al. (2011) Keratin gene mutations in disorders of human skin and its appendages. *Arch Biochem Biophys* 508(2): 123-137.
29. Freed-Pastor WA, Prives C (2012) Mutant p53: one name many proteins. *Genes Dev* 26(12): 1268-1286.
30. Yang L, Jiang Y, Shi L, Zhong D, Li Y, et al. (2020) AMPK: potential target for Alzheimer's disease. *Curr Protein and Peptide Sci* 21(1): 66-77.
31. Caughey B, Lansbury PT (2003) Protofibrils, pores, fibrils and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* 26: 267-298.

