Protein folding problem and the levinthal paradox biology essay

Science, Biology

\n[toc title="Table of Contents"]\n

\n \t

\n[/toc]\n \n

Abstract

Proteins fold rapidly in to their native three dimensional conformations in a manner which is favored both kinetically and thermodynamically. Despite of huge efforts, investments and claims made in the field of protein folding, the queries related to the general principles governing the folding, speed of folding and structure prediction based on primary sequence information still remains to be answered and hence the protein folding problem continues to prevail in the field. However, the forces and interactions involved in getting the protein from an unfolded state to a folded conformation are known and have been studied in great detail. Amongst numerous possible conformations, the one, in which the free energy of the system is least, is the native structure. The native structure of a protein in its microenvironment (temperature, pH, molecular crowding, solvent, presence of metal ions etc.)

is largely ascertained by atomic level interactions and therefore by its primary sequence. The folding and concomitantly the stability of the native structure is contributed by factors like hydrogen bonding, van der Waals interactions, electrostatic interactions, hydrophobic effect, presence of chaperones and overall by the microenvironment of the protein itself. Core packing is shown to be critical for the structural stability and function of proteins. The process of protein folding is largely pushed by the free energy of the structure that is acquired in advancing to the native structure. Several mechanisms like chaperone mediated, nucleation-condensation, folding co - operativity, hierarchic mechanism etc have been shown to exist for protein folding. Therefore, proteins seem to reach their folded state by diverse pathways. This review mainly touches upon the forces and interactions involved in protein folding and complexity associated with the folding process. It also talks about the methods employed to study this complex phenomena, technical challenges in the current scenario, recent progress made and the future prospects in the field.

Introduction

Proteins have the remarkable ability to fold rapidly in to their native conformation following synthesis in the cellular environment. The primary sequence of the protein dictates its three dimensional structure. But how this entire phenomenon happens is not clear and hence the term protein folding problem exists. The history of protein folding problem dates back to more than five decades. The field came in to highlight in 1961 when Christian Anfinsen demonstrated the refolding of ribonuclease to an active

conformation after denaturation (Anfinsen C, 1961). Since then, researchers continue to contribute substantially to this major challenge in biology- the protein folding problem. In 1968, Cyrus Levinthal came up with a paradox that if a protein folds by sampling all the possible conformations accessible to it, it would require astronomical amount of time even if each individual conformation is sampled at a nanosecond scale (Levinthal C, 1968). This is known as Levinthal paradox and it is described in detail in the next section. In 1976, Honig et al. speculated that proteins fold in a hierarchic manner in which formation of local structures happens first, followed by the interactions between these local structures to yield the native structure (Honig et al., 1976). The concept of folding on a multidimensional energy funnel was introduced by Dill and Chan in 1997 (Dill et al., 1997). This so-called ‘ new view’ of the protein folding considers energy landscapes and folding funnels instead of the ‘ folding pathways’. In the last decade, computational power has been used extensively to understand the folding dynamics. Presently, Baker’s group and Vijay Pande’s group rely on distributed supercomputers. Pande’s Folding@home and Baker’s Rosetta@home programs have made a phenomenal impact in the field of protein folding by speeding up predictions up to a hundred thousand-fold (News, August 2008, Science). Christian Anfinsen also proposed a hypothesis in relation to the thermodynamics of protein folding. This concept, occasionally called the ‘ thermodynamic hypothesis’ proposes that the native structure of the protein in its microenvironment is the one in which the Gibbs free energy of the system is minimum. This means that the three dimensional configuration of the protein is unique, stable in the environment and is accessible kinetically. Thus, the

native structure is thought to be at global minimum free energy as compared to any other state it can take on the same time scale. This, in turn, implies that in a specified environment, the functional state of the protein is determined by the interactions between atoms and therefore by the primary structure of the protein itself. Proteins are classic examples of molecules that represent a structure-function correlation at the molecular level. The native structure of the protein achieved upon folding is critical for its biological function. The three dimensional structure of the protein determines the mechanism by which it functions. The correct folding of proteins is not only important for performing its biological functioning but also important for maintaining cellular homeostasis and eventually for physiological well being. Misfolded or unfolded proteins are known to aggregate in the cell and this leads to many diseases like neurodegenerative diseases, prion related illness etc. This will be discussed in later half of the review.

Protein folding problem and the Levinthal paradox

The entire problem of protein folding narrows down to the issues of the speed of protein folding, determination of the physical code by which the primary sequence dictates the native structure and the ability to devise computer programs that can predict the functional structure of the protein with the sequence information (Dill K. A and MacCallum J. L, 2012). The most intimidating part of the problem is predicting the structure of the unknown proteins. The entire process of protein folding is very complex and involves a large number of interactions and factors that play major role in the process. The complexity is not only due to the primary structure of the protein, but

also due to macromolecular crowding and the spatially non-homogenous environment of the cell. Studying protein folding in vitrobecomes difficult as the cellular conditions like compartmentalization, molecular crowding, hindered diffusion etc. cannot be met in vitro (Gershenson and Gierasch, 2011). speed of prot folding???? Even if a protein samples one different conformation in a fraction of second, it would require time as huge as the age of universe or even folds of magnitude more than that to fold in to the right conformation (Levinthal C, 1968). But proteins fold rapidly in to their native structures despite of huge number of conformations it can access. This paradox was first brought up by Cyrus Levinthal and since then it has been referred to as Levinthal paradox (Levinthal C, 1968). This gave rise to the fact that proteins do not achieve their native state randomly. For a protein with n residues, the total number of distinct structural conformations estimated is 10n. This turns out to be an astronomical number even for a small protein and therefore a protein cannot reach its native conformation in a purely random basis (Levinthal C, 1968). The Levinthal paradox does not hold true to be called a paradox because of its assumption that proteins fold in to their native structure randomly and there are no favorable interactions that gradually increase as the protein folds towards its native structure. It has been established since long that interactions govern the conformational sampling of the proteins and this makes it far from random (Kauzmann, 1959). In addition to this, the Levinthal paradox breaks down in light of thermodynamic parameters. Levinthal hypothesis requires significant decrease in entropy without a gain in the enthalpy, i. e. very high activation energy. This does not match with the experimental observations wherein

folding is associated with low activation free energy (Wedmeyer and Scheraga, 2001).

Factors affecting protein folding:

Proteins have distinct structural features with a defined internal architecture and form compact globular structures. A large number of forces and interactions contribute to the folding of a protein thereby leading to a stable three dimensional structure. Hydrogen bonding, Van der Waals interactions, electrostatic interactions, hydrophobic effect and the microenvironment contribute to the free energy of folding. Hydrogen bonding is a dominant force in protein folding as amino acids are dipolar and hence can form hydrogen bonds (Dill, 1990). Also the secondary structures like α-helix and β-sheets exhibit extensive hydrogen bonding and promotes organization within the interior of the compact protein (Dill K. A and MacCallum J. L, 2012)(Dill, 1990). An increase in the number of hydrogen bonds has significant correlation to bettered thermo stability of the protein (Vogt et al., 1997). Van der Waals forces are short ranged interactions between the tightly packed atoms in a protein that stabilize the structure locally (Dill, 1990). Van der Waals interactions are weaker interactions but the total of these interactions in a protein molecule adds up to huge amount and thus contributes to the stability of the native structure. Electrostatic interactions, another important contributor to protein folding, are charge dependent long- ranged interactions. Electrostatic interactions depend on the pH and the ionic strength of solution and involve charged amino acids. The overall charge on the protein is determined by the pH whereas the salt concentration dictates

the degree of interactions between the charged residues. The formation of salt bridge or ion pairing happens when the side chains of oppositely charged amino acids are in close spatial proximity. Unlike the classical mechanism which postulates a decrease in protein stability with an increase in charge, the ion pairing effect might increase the stability of the protein (Dill, 1990) (Vogt et al., 1997). However, ion pairing is not the predominant force in protein folding. One of the major factors leading to the folding of a protein in to a compact well packed structure is the hydrophobic effect. The non polar hydrophobic amino acids lie in the interior of the protein, isolated from the polar aqueous environment whereas the polar residues form the surface of the folded protein and interact with the aqueous milieu (Dill, 1990). It is not a force on its own; rather it is a consequence of disrupting the extensive hydrogen bonding of the surrounding water molecules. The hydrophobic amino acids getting sequestered in the core of the protein and avoiding contacts with the aqueous environment imparts thermodynamic stability to the native structure. Hydrophobic effect has been demonstrated to be a dominant factor in the initial stages of protein folding. Computational results show the occurrence of hydrophobic collapse in conjugation with the segregation of the hydrophobic and hydrophilic residues in the earliest stage of folding (Lin et al., 2012). Recent experiments support the long standing hypothesis that the folding initiates from hydrophobic interactions (Dyson, 2006). Many proteins have multiple metastable states that vary in their hydrogen bonding patterns and hydrophobic contacts (Bywater, 2012). This also highlights the importance of these forces in protein folding. The microenvironment of the protein plays a crucial role in shaping the folding

pathways (Anfinsen C, 1961). The native structure of the protein is achieved in its milieu of a particular pH, temperature, ionic strength, presence of other macromolecules, presence of metal ions and co-factors etc. Change of pH can affect the electrostatic interactions in the protein, thereby affecting protein folding. Folding is more favored at physiological temperature, the efficiency of folding decrease in slight changes in temperature and proteins don’t fold at extremes of temperature. The above mentioned dominant forces in protein folding represent just one side of the coin. It is likely that there are ‘ opposing forces’ of magnitude matching to that of the above mentioned forces (Dill K. A., 1990) but this has been neglected till date. It is likely that there is a fine balance between the dominant forces and the opposing forces, which might be leading to the native, stable and yet flexible conformation of the protein. The opposing forces could simply be intrinsic stearic hindrance, repulsive forces between amino acids of the same charge lying in close spatial proximity, etc. Thus, it is expected that the magnitudes of the driving and opposing forces are approximately the same and it is the balance between these two force categories that stabilize the three dimensional structure of the protein. Molecular dynamics simulations indicate that there must be a fine balance between the hydrophobicity and hydrogen bonding for proteins to fold to a correct structure (Fitzpatrick et al., 2011). Thus, having involved a variety of forces and interactions, protein folding pathway is a complex phenomenon.

Folding mechanisms- a brief overview:

Decades of research has gone in to propose and prove various mechanisms underlying protein folding. A few of the mechanisms are classical nucleation, nucleation-condensation, hydrophobic collapse, nucleation-condensation, cooperative, hierarchic and chaperone mediated mechanisms. Both experiments and theory indicate the existence of nucleation mechanism in protein folding. Classic ‘ nucleation mechanism’ proposes that certain proximal residues would interact to form secondary structure i. e. a strong localized nucleus that would serve as a template for the folding of the remainder of the protein in a stepwise manner (Fresht, 1997). On the other hand, the ‘ hydrophobic collapse’ mechanism postulates the folding of a protein because of rapid collapse around the hydrophobic regions thereby leading to compaction and folding of the protein in a confined space. This decreases the conformational search of the protein to the native state (Fresht, 1997) (Daggett & Fersht, 2003). This mechanism supports the existence of folding intermediates whereas the nucleation mechanism does not (Fresht, 1997); therefore, the nucleation mechanisms are no longer believed to exist. Later, the concept of ‘ nucleation-condensation’ was hypothesized. This model combines unifies the concepts of both nucleation and hydrophobic collapse and proposes that weak nucleation happens in one region of the protein which is then stabilized by interactions from elsewhere in the same protein giving rise to a broadened nucleus (Fresht, 1997). The nucleus consolidates as folding proceeds (Fresht, 1997). The nucleation-condensation mechanism seems to explain the folding of small proteins. Larger proteins are speculated to fold as modules that can independently

fold by nucleation-condensation mechanism and the finally achieve the native conformation. One of the hallmarks of protein folding is Cooperativity. Protein folding is highly cooperative in that once the folding initiates, further folding is more favored and this speeds up the entire process. Mostly, all the proteins are either in fully folded or fully unfolded states and do not stabilize in a state which is intermediate between the two. A small perturbation in one region hampers the stability of the entire protein (Baldwin and Rose, 1999). This means that the residues interact with each other in a cooperative manner to form a stable native structure. ‘ HZ’ (hydrophobic zipper) hypothesis has been proposed (Dill et al., 1993) which postulates that other contacts are brought in to close spatial proximity by hydrophobic contacts, which then further stimulate the process and zip up other contacts. Computational studies also show that unfolded proteins follow a two state folding transition pathway to the native conformation and it exhibits Cooperativity (Hao et al., 1998). ‘ Hierarchic mechanism’ (Baldwin and Rose, 1999) proposes that proteins fold hierarchically and have populated intermediates. This means that folding initiates locally i. e. the secondary structure is determined first, and then the tertiary conformation is achieved. The local structures are marginal in terms of stability. However, interactions between these local structures generate intermediates which lead to the three dimensional structure. Molecular dynamics simulations also reveal a hierarchical mechanism wherein the α-helices are locally formed before the entire native structure is achieved (Scott et al., 2006). Molecular chaperones (e. g. Hsp70) and chaperonins (e. g. GroEL) are key players involved in the process of protein folding. Chaperones perform the remarkable role of

upholding the translating polypeptide chains or the nascent unfolded polypeptides in a folding-competent state whereas chaperonins have a large central cavity which prevents the aggregation of unfolded polypeptides (Netzer and Hartl, 1998). Molecular chaperones bind to the hydrophobic residues of the nascent polypeptide, thereby maintaining the folding competence of the polypeptide (reviewed in detail in ref. 23). Consequently, the tendency of the polypeptide to aggregate is also reduced to a great extent. Chaperones, in a way, prevent any off-pathway folding reaction and hence favor the folding of the polypeptide in a correct manner. Chaperonins, on the other hand, are cylindrical complexes that help in folding of polypeptides that are sensitive to aggregation (Netzer and Hartl, 1998). A nascent polypeptide enters the central cavity (" Anfinsen’s cage") of the chaperonin, folds in the non-permissive folding conditions of the cavity and therefore prevents aggregation. The mechanism of functioning of GroEL is well established (Hartl, F. U., 1996). Recently, it has been shown that through passive action of preventing multimolecular aggregation, the chaperonins increase the rate of protein folding (Apetri el al., 2008). Molecular dynamics simulations illustrate that rate of protein folding increases as a result of confinement in chaperonin mediated protein folding (Baumketner et al., 2003).

Current scenario and challenges

Five decades on, the field of protein folding seems to have made a substantial progress in a gradual manner rather than in big steps. Anfinsen’s paradigm " Sequence determines structure" still remains undefeated and

upheld. The field now has answers to Levinthal paradox and clearly this paradox breaks down because of the reasons mentioned above. Recent advances in experimental techniques have assisted in studying protein folding in great detail and this has led to better connection between experimental and theoretical studies. As a consequence, some basic and longstanding questions in protein folding field have been answered and a better understanding of the basic concepts underlying the folding mechanisms has been achieved. The growth of Protein Data Bank (PDB) which has more than 80, 000 structures deposited (till April, 2013) is an indirect evidence of the growth in the field as well as a great support for computational techniques in protein folding. The current state-of-art understanding of the field is the concept of funnel-shaped energy landscape. Funnel-shaped energy landscape hypothesizes that multiple copies of the same protein fold via different routes to give the same native structure ultimately (Dill, 2012). It is possible that some routes are more populated as compared to the others. As of now, the field has reached a state where we are aware of the basic mechanisms that lead to folding and the fundamental interactions involved in the native structure of a protein. Atomic level understanding of the field has been boosted by remarkable advances in computational methods. The field is heading more towards computational studies, especially the use of the highly powerful and fast supercomputers. Blue Gene supercomputer, developed by IBM, is largely dedicated to understand the folding of proteins. This gave insights about protein folding at atomic resolution which has taken the entire field to a higher level where researchers have started to look protein folding using statistical mechanics

and other physical parameters. Also the unique concept of distributed-grid computing has driven significant advances in the field of protein folding. Folding@home, software designed by V. S. Pande at Stanford is an example of distributed-grid computing in which simulations of protein systems are carried out by computer users all round the world. Recent advances in experimentally understanding protein folding involves methods like temperature-jump perturbation technique coupled with fast relaxation imaging (FReI) that can help to investigate folding landscape of protein of interest in vivo (Ebbinghaus et al., 2010). Also a novel combination of hydrogen exchange and mass spectrometry method has been used to check the stability of proteins in vivo (Ghaemmaghami et al., 2001). Even though the field has progressed considerably, it is still being challenged by the technical limitations in studying the folding pathways. Till date, we do not have a technique that can experimentally track the folding pathway along with its intermediates. Proteins fold rapidly and no technique can actually keep a track of such a fast event, all we see is the unfolded and the final folded state of the protein. Another important reason for the field being unable to draw an exact picture of protein folding pathway is that the in vitro experiments are performed under dilute conditions which are not even closer in mimicking the highly crowded cellular environment. Thus the results obtained from in vitro experiments might be highly misleading unless macromolecular crowding is accounted for in the experiments. Macromolecular crowding in the cell might have two opposite roles. First, assisting the protein to fold by limiting it to a confined area where the protein does not have much to do except than to fold. Second, it might

retard the entire process of folding by not providing the correct environment to fold. In vitro experimental conditions do not mimic the inhomogeneous interior of the cell and this is another major limitation in the field of protein folding. Apart from the limitations in experimental protein folding, computational protein folding also faces a lot of challenges. Firstly, even with the fastest computers, folding a small polypeptide chain takes considerable amount of time. Folding a small polypeptide would take approximately an year or so. Therefore, computational speed limitation is a majorchallenge in studying protein folding. Also validating computational results in physiological conditions is a daunting task. Peter Wolynes, an expert in the field, claims that the protein folding problem is now solved where as his rival David Baker considers such claims to be ‘ dangerous’ and might under estimate the interest in the field. Ken Dill and other pioneers in the field all agree that the field has made significant progress. " We have not been able to transfer our conceptual understanding into [a] prediction of how specific amino acid sequences will fold" says Andrej Sali, another expert in the field (adapted from news article- Problem solved, Vol 321, Science, 2008). Despite of huge efforts, investments and claims made, the queries related to the general principles governing the folding, speed of folding and structure prediction based on primary sequence information still remains to be answered and hence the protein folding problem continues to prevail in the field.

Misfolding of proteins can have biological consequences

Protein folding is important to arrive at the three dimensional native conformation and the native structure of the protein in critical for its function. A partially folded or misfolded protein cannot perform the function that it is supposed to do; rather it becomes a potential threat to cellular homeostasis. Misfolded or unfolded proteins have increased chances of aggregating in the cell and these cellular aggregates have been shown to cause various disorders. Such polypeptides deposit in the form of amyloid fibrils and plaques, a feature observed in a number of neurodegenerative conditions like Alzheimer’s disease, Parkinson’s disease, prion diseases etc (Stefani and Dobson, 2003). This highlights the importance of correct folding of proteins for the maintenance of cellular homeostasis.

Unaddressed mysteries of protein folding

The protein folding problem still remains to be a black box with numerous issues still remaining to be addressed. A holistic view with aggressive multidisciplinary research would be required to find solutions to the open questions in the field. The major unaddressed issues in the field are: (a) determination of the physical code that govern protein folding, (b) prediction of the native conformation of the protein on the basis of its primary structure, (c) understanding the folding behavior of proteins in compartmentalized and highly crowded cellular environment rather than the dilute homogenous in vitro environment, (d) the propensity of proteins to aggregate in cell, (e) detailed understanding of the folding intermediates, their nature and the folding pathways, (f) connection between protein

aggregation and folding diseases, and (g) development of novel techniques for analyzing the folding dynamics. Protein folding is now a well established field of active research in interdisciplinary biological sciences and the future of the field remains as promising as the past and the present. Much of the hope relies on the emerging computational power. Coming up with an apparently reasonable model that overcomes the Levinthal paradox is not what protein folding studies are focusing on. The ultimate and the larger goal is to understand the general rules that govern the folding of any given protein, to predict the native structure of the protein based on its sequence, and to develop high resolution techniques and faster computational algorithms to enhance the current understanding and to unravel the unsolved mysteries of one of the biggest problems in biology –the protein folding problem.