Journal of Glycomics & Lipidomics

Protein Folding and Misfolding: A Perspective from Theory

Naeem A[1]*, Khan TA[2] and Fazili NA[1]

[1]Department of Biochemistry, Faculty of Life Sciences, AMU, Aligarh 202002, India
[2]Applied Biotechnology Department, Sur college of Applied Sciences, P.O. Box 484, Sur, 411, Oman

Abstract

The understanding of folding of proteins into their compact three-dimensional structures, example of complex biological self-assembly process, will provide an insight into the way in which evolutionary selection has influenced the properties of a molecular system for functional advantage. Once regarded as a grand challenge, protein folding has seen much progress in recent years. Protein folding pathways are of great interest not only in themselves, but also because understanding them is important for both protein structure predictions and for de novo protein design. Protein misfolding is a ubiquitous phenomenon associated with a wide range of diseases. Aggregation of misfolded proteins that escape the cellular quality-control mechanisms is a common feature of a wide range of highly debilitating and increasingly prevalent diseases. We hope that this review will stimulate further research in this area and catalyze increased collaboration at the interface of chemistry and biology to decipher the mechanisms and roles of protein folding, misfolding and aggregation in the fields of health and disease.

Keywords: Native state; Unfolded state; Molten globule; Folding code; Aggregation

Introduction

In appropriate physiological milieu proteins spontaneously fold into their functional three-dimensional structures. The amino acid sequences of functional proteins contain all the information necessary to specify the folds [1,2]. The manner in which a newly synthesized chain of amino acids transforms itself into a perfectly folded protein depends both on the intrinsic properties of the amino-acid sequence and on multiple contributing influences from the crowded cellular milieu. The wide variety of highly specific structures that result from protein folding and that bring key functional groups into close proximity has enabled living systems to develop astonishing diversity and selectivity in their underlying chemical processes. In addition to generating biological activity, we now know that folding is coupled to many other biological processes, including the trafficking of molecules to specific cellular locations and the regulation of cellular growth and differentiation [3,4].

Add to it, correctly folded proteins have long-term stability in crowded biological environments and are able to interact selectively with their natural partners. It is therefore not surprising that the failure of proteins to fold correctly is the origin of a wide variety of pathological conditions [5]. Some concepts, such as energy landscape, Gibbs energy landscape and co-operativity frequently used in the theory of protein folding are examined exactly in one-dimensional systems. It is shown that much of the confusion that exists regarding these, and other concepts arise from the misinterpretation of Anfinsen's thermodynamic hypothesis [6].

Moreover, proteins are involved in virtually every biological process and their functions range from catalysis of chemical reactions to maintenance of the electrochemical potential across cell membranes. The molecular conformation of proteins is sensitive to the nature of the aqueous environment [7]. They are synthesized on ribosomes as linear chains of amino acids in a specific order from information encoded within the cellular DNA. To function, it is necessary for these chains to fold into the unique native three dimensional structures that are characteristic for each protein (Figure 1). This involves a complex molecular recognition phenomenon that depends on the co-operative action of many relatively weak non-bonding interactions. As the number of possible conformations for a polypeptide chain is astronomically large, a systematic search for the native (lowest energy) structure would require an almost infinite length of time. Recently, significant progress has been made towards solving this paradox and

understanding the mechanism of folding. This has come about through advances in experimental strategies for following the folding reactions of proteins in the laboratory with biophysical techniques, and through progress in theoretical approaches that simulate the folding process with simplified models [8,9].

Protein folding is a problem of great importance in both life sciences and biotechnology industries. A very large number of distinct conformations exist for the polypeptide chain of which a protein molecule is composed. The protein spends most of its time in the native conformation, which spans only an infinitesimal fraction of the entire configuration space [10]. The general perception has been that the protein folding problem is a grand challenge that will require many supercomputer years to solve. During the past few years, there has been a great increase in the level of interest in protein folding. This is due in part to the challenge of the human genome project and in part to the development of experimental methods that provide more details about the folding process. One conclusion from the measurements is that an essential part of the folding process, the search for an ordered globule with many attributes of the native structures, is composed within the dead time [11]. Much less is known about the features of the potential surface governing the non-native portion of configurational space involved in protein folding. This includes a wide range of structures that may differ by tens of angstroms and be separated by significant energy barriers [12].
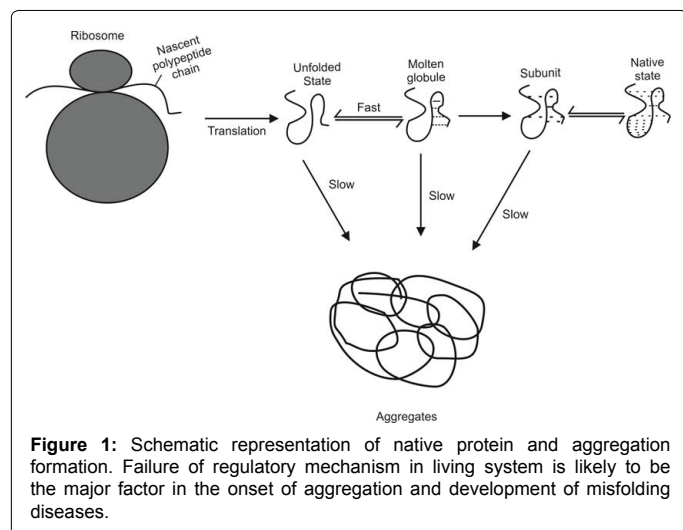
In addition, when a denatured polypeptide chain is placed under refolding conditions, the van der waals and electrostatic interactions within the protein and between the protein and solvent stabilize the native state. However, the greater stability of the native state relative to the denatured state does not by itself explain how the polypeptide chain

Figure 1: Schematic representation of native protein and aggregation formation. Failure of regulatory mechanism in living system is likely to be the major factor in the onset of aggregation and development of misfolding diseases.

finds the former (a single state out of an astronomically large number of denatured configurations) starting from the latter. In one scenario of protein folding, the polypeptide chain collapses rapidly to a compact globule [13]. Although the most significant factor leading to such an early collapse is the burial of hydrophobic groups and the nature of the collapse depends on the heterogeneity of the stabilizing interactions. The presence of ions can also stabilize or destabilize (denature) protein secondary structure [7]. Folding is a progression in which non-native and native contacts, some of which may be particularly important, stabilize native-like features of the structure. Another essential aspect of finding the correct fold is the need to avoid highly destabilizing situations, such as unbalanced charged residues in the protein interior [8,14].

Recombinant DNA technology facilitates the design and expression of various proteins at high expression rates and concentrations. "However, the production of recombinant protein is often hindered by the refolding process in which the expressed protein attains its native-like three-dimensional conformation that enables its designated biological functions". Misfolding and aggregation are major problems in protein refolding *in vitro* [15]. Macromolecular crowding is recognized as an important factor influencing folding and conformational dynamics of proteins [16]. Direct dilution of the protein dissolved in denatured buffer containing either 8 M urea or 6 M Guanidine HCl remains the most widely used method for protein refolding at different scales. Protein folding is so complex that even the approaches to be used for its characterization are not fully defined [17]. Anfinsen established that small proteins could spontaneously refold from their denatured states and that the primary sequence of a protein dictates its tertiary structure. In addition, he described that refolding is the art of establishing a kinetic partition that drives the protein toward the native conformation in which it attains its thermodynamically most stable form [18]. The integrated use of protein engineering, molecular simulation, and bioinformatics tools has enabled understanding and visualizing the folding pathways for several proteins at atomic resolution [19]. These studies have shed light on various aspects of the process and will be helpful in engineering the pathways for protein refolding *in vitro* [20]. Protein refolding is the art of establishing a kinetic partition that drives the protein toward the native conformation in which it attains its thermodynamically most stable form, as described by Anfinsen [18]. Proteins that show reversible two-state folding–unfolding transitions turned out to be a gift of natural selection. Focusing on these simple systems helped researchers to uncover general principles regarding the origins of co-operativity in protein folding thermodynamics and kinetics [21].

Moreover, globular proteins are synthesized as linear chains of amino acids. To carry out their functions, they must fold rapidly and reliably to a specific structure designed by evolution for the particular task. Although the folding process in a cell involves a range of catalytic and control systems [22], the information for folding is contained in the amino acid sequence for many, if not all, proteins [18]. A protein sequence must satisfy two requirements: one thermodynamic and the other are being kinetic [13]. The thermodynamic requirement is that the molecules adopt a unique folded conformation (the native state) which is stable under physiological conditions. The kinetic requirement is that the denatured polypeptide chain can fold into the native conformation within a reasonable time. Protein folding is now at the stage where theory and experiment can together make rapid progress toward an understanding of this complex process. Toward this end, a framework must be developed for interpreting the results of the new experimental techniques and for probing specific features of the folding reaction [8,23].

Stable conformational states of proteins

Different protein conformations differ only in the angle of rotation about the bonds of the backbone and amino acid side chains, except that they may also differ in covalent disulphide bonds, which are unique.

Native state

In contrast, homologous proteins invariably have essentially the same folded conformation, even if their amino acid similarities are minimal [24,25]. The most conspicuously similar aspects of homologous structures are the general conservation of the non-polar character of the side-chains that comprise the folded interior, plus the general prevalence of hydrophilic side-chains at the surface [26]. Proteins have been, found to be surprisingly adaptive to mutations that would be expected to be disruptive, but the hydrophobic core seems to be the most critical aspect for stability of the normal folded state [27]. Folded proteins demonstrate varying degrees of flexibility [28], which is of direct relevance to protein folding, in that it reflects the free energy constraints on unfolding and refolding. Flexibility is greatest at the protein surface, where some side chains and a few loops have alternative conformations or no particular conformation that is energetically preferred [29]. Although flexibility is least in the interior, even their side-chain rotations occur, and most tyrosine and phenylalanine side-chain aromatic rings are flipping by 180° on the millisecond time-scale [30].

For some proteins, probes for different aspects of structure reveal very different kinetic behaviour during folding. One of the earliest indications of this was that for many proteins the extrinsic fluorescence of a dye (8-anilino-1-naphthalinesulfonic acid, ANS) added to the refolding solution and develops prior to the intrinsic fluorescence of the aromatic side chains [31]. As the former is indicative of the formation of clusters of hydrophobic residues, and the latter of native-like packing of side chains, this suggests that collapse to a relatively disorganized globule occurs before the formation of native-like structure [8,32]. Osmolytes produced under stress in animal and plant systems have been shown to increase thermal stability of the native state of a number of proteins as well as induce the formation of molten globule in acid denatured states and compact conformations in natively unfolded proteins [33].

Unfolded state

The ideal unfolded protein is the random coil, in which the rotation angle about each bond of the backbone and side-chains is independent of that of bonds distant in the sequence, and where all conformations have comparable free energies, except when atoms of the polypeptide chain come into close proximity. The protein folding process begins from the unfolded state and progresses to the native or misfolded states via various kinds of folding intermediates (Figure 1) [34]. Understanding the mechanism of small proteins folding requires characterization of their starting unfolded states and any partially unfolded states populated during folding [35]. Also, each molecule in a typical experimental sample of a fully unfolded protein (likely to contain no more than 1018 molecules) will probably have a unique conformation at any instant of time. Consequently, the initial stages of folding must be nearly random, but the native conformation is unlikely to be found by a totally random process. There is a wide variety of evidence, however, suggesting that unfolded proteins are not true random coils under other conditions, such as extremes of pH or temperature in the absence of denaturants [36]. Unveiling the structural and dynamic properties of these states, particularly the unfolded states and folding intermediates, at an atomic level is crucial for understanding protein folding pathways [37].

In addition, several theoretical and computational studies have demonstrated that specific interactions of denatured proteins play important roles in biasing the conformation toward the native state, thereby affecting the folding pathway [38]. Molecular simulations have been used to illustrate the unfolded protein structure and its radius of gyration at the molecular level. These would provide a better understanding of the controversial properties of denatured proteins, the random coil scaling of their sizes, and the presence of residual native structures [39]. Molecular dynamics simulations of the unfolding of cytochrome c' at 498 K shows that the reverse turns persist in the unfolded state. Thus, the portions of the primary structure of human cytochrome c set up the topology of cytochrome c in the denatured state, predisposing the protein to fold efficiently to its native structure [40]. The results suggest that denatured proteins follow the random coil scaling sizes with statistical validity and retain residual secondary structures similar to those found in their native states. These computational studies provide conceptual reconciliation between two seemingly mutually exclusive views of protein unfolded states. In physical terms, the existence of residual native secondary structure indicates the possibility of avoiding any excessive entropic energy barrier to the formation of both secondary and tertiary structures, which would consequently accelerate the folding rate [22,41].

Molten globule state

A variety of proteins have been observed under certain conditions to exist in stable conformations that are neither fully folded nor fully unfolded. These conformations have sufficient similarities to suggest that they are different manifestations of a third stable conformational state [42]. Conformational transition from the native to the molten-globule form proceeds in a stepwise manner involving a burst-phase with a submillisecond conformational change followed by biphasic slower conformational reorganizations on the millisecond time scale leading to the final molten-globule state [43]. Due to the elusive nature of protein folding, the identification and characterization of intermediates is an extremely difficult task. The overall dimensions of the polypeptide chain are much less than those of a random coil and only marginally greater than those of the fully folded state. The average content of secondary structure of molten globule is similar to that of the folded state. The interior side-chains are in homogeneous surroundings, in contrast to the asymmetric environments they have in the fully folded state. Many interior amide groups exchange hydrogen atoms with the solvent more rapidly than in the folded state, but more slowly than in the fully unfolded state. Its enthalpy is very nearly the same as that of the fully unfolded state, substantially different from that of the native state. Inter-conversions with the fully unfolded state are rapid and non-co-operative, but slow and co-operative with the fully folded state [44].

Moreover, the traditional view of protein folding considers "intermediates" to be partially unfolded conformations that are stable enough to be detected. It has been found that partially unfolded intermediates are present in the protein folding pathways of even conventionally observed two-state small proteins, i.e. those composed of less than 100 amino acids. To understand the stabilization, folding, and functional mechanisms of proteins, it is very important to understand the structural and thermodynamic properties of the molten globule state [45]. The use of single molecule methods such as fluorescence resonance energy transfer [46] and force spectroscopy, including optical tweezers and AFM [47], may enable the detection of folding intermediates. Various computational groups have made great efforts to explore protein folding intermediates and relative folding pathways. The simulation results agree well with those obtained from both traditional experimental NMR-based studies [48] and single molecule based detection techniques [49].

Using optical trap force spectroscopy, Elms *et al.* [50] investigated the response to force of the native and molten globule states of apomyoglobin along different pulling axes. Unlike natively folded proteins, the molten globule state of apomyoglobin is compliant (large distance to the transition state); this large compliance means that the molten globule is more deformable and the unfolding rate is more sensitive to force (the application of force or tension will have a more dramatic effect on the unfolding rate). Studies of Elms *et al.* [50] suggest that these are general properties of molten globules and could have important implications for mechanical processes in the cell.

Trapping of folding intermediates by disulphide bonds

Elucidating the mechanism of protein folding requires characterization of the initial, final and intermediate conformational states, plus the determination of steps by which they are inter-converted [51]. The ideal situation would be to control the rates of formation and breakage of hydrogen bonds, since every protein structure includes them. During folding, molecules with 1, 2, 3,... intra-molecular hydrogen bonds might accumulate kinetically; if they could be trapped and identified, a pathway could be defined in terms of hydrogen bonding. It is unfortunately not possible to trap hydrogen bonds, but disulphide bonds can be trapped, due to the reduction-oxidation nature of the covalent disulphide interaction between thiol groups [52]. The approach is only useful with proteins that unfold when their disulphides are broken; unfolding and refolding can then be controlled by varying just the intrinsic disulphide stability. There is no need to use denaturants, and the strengths of all other types of interactions that stabilize proteins can be kept constant.
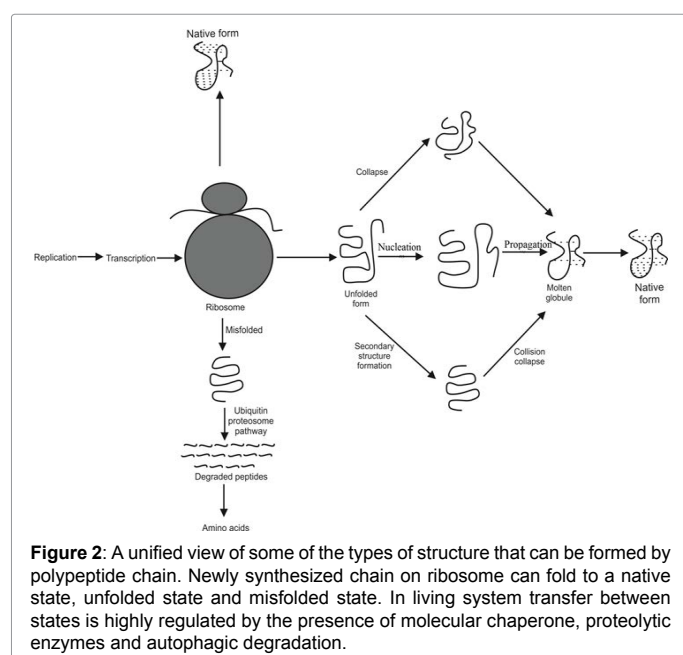
The folding code

From his now-famous experiments on ribonuclease, Anfinsen postulated that the native structure of a protein is thermodynamically the most stable structure; it depends only on the amino acid sequence and on the conditions of solution but not on the kinetic folding route. It became widely appreciated that the native structure does not depend

on whether the protein was synthesized biologically on a ribosome or with the help of chaperone molecules, or if instead, the protein was simply refolded as an isolated molecule in a test tube. Two powerful conclusions followed from Anfinsen's work [18]. First, it enabled the large research enterprise of *in vitro* protein folding that has come to understand native structures by experiments inside test tubes rather than inside cells. Second, the Anfinsen principle implies a sort of division of labor: Evolution can act to change an amino acid sequence, but the folding equilibrium and kinetics of a given sequence are then matters of physical chemistry [53].

A key idea was that the primary sequence encoded secondary structures, which then encoded tertiary structures [54]. Because native proteins are only 5-10 kcal/mol more stable than their denatured states, it is clear that no type of intermolecular force can be neglected in folding and structure prediction [55]. Folding is not likely to be dominated by electrostatic interactions among charged side chains because most proteins have relatively few charged residues; they are concentrated in high-dielectric regions on the protein surface. Protein stabilities tend to be independent of pH (near neutral) and salt concentration, and charge mutations typically lead to small effects on structure and stability. Hydrogen-bonding interactions are important, because essentially all possible hydrogen-bonding interactions are generally satisfied in native structures. Similarly, tight packing in proteins implies that Van der Waals interactions are important [56,57].

Moreover, there is considerable evidence that hydrophobic interactions must play a major role in protein folding. Proteins have hydrophobic cores; implying non-polar amino acids are driven to be sequestered from water. Model compound studies show 1-2 kcal/mol for transferring a hydrophobic side chain from water into oil-like media [58], and there are many of them. Proteins are readily denatured in non-polar solvents. Sequences are jumbled and retain only their correct hydrophobic and polar patterning fold to their expected native states [59], in the absence of efforts to design packing, charges, or hydrogen bonding. Hydrophobic and polar patterning also appears to be a key to encoding of amyloid-like fibril structures [60]. Secondary structures in proteins are substantially stabilized by the chain compactness, an indirect consequence of the hydrophobic force to collapse (Figure 2).

Figure 2: A unified view of some of the types of structure that can be formed by polypeptide chain. Newly synthesized chain on ribosome can fold to a native state, unfolded state and misfolded state. In living system transfer between states is highly regulated by the presence of molecular chaperone, proteolytic enzymes and autophagic degradation.

Like airport security lines, helical and sheet configurations are the only regular ways to pack a linear chain (of people or monomers) into a tight space.

General properties of protein folding

The native conformational states of proteins may usually be unfolded reversibly by adding denaturants, increasing or decreasing the temperature, varying the pH, applying high pressures, or cleaving disulphide bonds. At equilibrium, unfolding transitions of single-domain proteins are usually two state: Native and Unfolded.

Stability of the unfolded state

Understanding the physical basis of stability of the folded state is crucial for understanding how such a conformation can be acquired. Natural proteins appear to require some flexibility for their function, or possibly to be able to fold into the native conformation quickly, both of which would be hampered by too stable a final folded conformation. The heat capacity of the unfolded state is significantly greater than that of the folded state. The usual interpretation is that the heat capacity difference results primarily from the temperature-dependent ordering of water molecules around the non-polar portions of the protein molecules, more of which are solvent accessible in the unfolded state [61], although other factors may contribute [62].

The other types of interactions present within folded proteins, such as hydrogen bonds, van der Waals' interactions and electrostatic interactions have traditionally been assumed to provide no net contribution to the overall stability of the folded state, because comparable interactions should be made between the unfolded state and the solvent. This conclusion, however, neglects the intra-molecular nature of the interactions within the folded state, as opposed to the intermolecular interactions between solvent and protein [63]. For example, most hydrogen bonds within water [64] and between protein and water are usually present only a fraction of the time, whereas those within folded proteins are present essentially all of the time; the latter should consequently have the more negative enthalpy. Van der Waals' interactions within the close-packed protein interior should be substantially greater than those between the protein and the solvent; they should have correspondingly lower enthalpies, analogous to the enthalpy of fusion when liquids crystallize [65]. It therefore seems reasonable to conclude that most of the interactions within the folded protein are more favourable energetically, in both enthalpy and free energy, than the corresponding interactions of the unfolded state. They should therefore contribute to the net stability of the folded state, and it is not surprising those hydrogen bonds and salt-bridges have been found to do so [66]. Nevertheless, the hydrophobic interaction is probably the major stabilizing factor [67-69].

Co-operativity of folding

The two-state nature of protein folding transitions indicates that folding is a co-operative process. Little happens to the fully folded state prior to complete unfolding; e.g. it is not detectably perturbed by varying the temperature [70]. Effects of denaturants are observed [71] but may arise because of direct interactions with the protein. Once unfolding of a domain is initiated, it proceeds to completion. The interactions stabilizing the folded conformation must be co-operative; breaking one or more of the interactions must destabilize the others so that the free energy increases and the folded conformation become unstable. Co-operativity of denaturant-induced unfolding is often inferred if it occurs over a limited range of denaturant concentration, to give a sigmoidal unfolding curve. This, however, only indicates that

the equilibrium constant for folding in the absence of denaturant is sufficiently large and that the proportion of unfolded molecules is negligible initially. There is no equivalent of the Henderson-Hasselbach equation to predict the properties expected for disruption of a single interaction, so such a sigmoidal unfolding curve is not necessarily evidence for co-operativity. Under stress conditions osmolytes stabilize co-operative protein structure relative to non-cooperative ensemble. These findings have implications towards the structure formation, folding and stability of proteins produced under stress in cellular systems [27].

There are probably two major reasons for the co-operativity of folding transitions. The first concerns unfavourable interactions in the partially-folded states that are not present in either the fully folded or unfolded states. The second reason for co-operativity is likely to be the entropic co-operativity of the folded state, which is necessary to account for its stabilization by intrinsically weak interactions [57]. The co-operative nature of protein substructure and internal motion is a critical aspect of their functional competence about which little is known experimentally [72]. The fully native structure is only acquired when all the native-like interactions have been formed both within and between the domains; this happens in a final co-operative folding step when all the side chains become locked in their unique close-packed arrangement and water is excluded from the protein core [73]. This co-operativity arises from the simultaneous presence of many interactions within a single conformation; their total contribution to stability may be much greater than that of the sum of the individual interactions (Figure 2) [74].

Protein Stability and Intermolecular Interactions

Proteins exhibit marginal stabilities equivalent to only a small number of weak intermolecular interactions [75]. Average values for the Gibbs free energy of stabilization (DG°_{stab}) of medium size globular proteins are on the order of 50 kJ mol^{-1} [76]. With regard to the intrinsic effects, stabilization may involve all levels of the hierarchy of protein structure, local packing of the polypeptide chain, secondary and super secondary structural elements, domains and subunits [77,78]. Hydrogen bonds are favoured at low temperature and become weaker as the temperature is increased. With regard to the stability of protein fragments, it has been known for a long time that proteins are co-operative structures showing mutual stabilization of structural elements. In addition to the sequence-encoded increments discussed so far, extrinsic factors such as ions, cofactors, metabolites, compatible solutes and covalent conjugates may contribute to protein stability [79]. The driving forces responsible for protein folding and protein stabilization are the same, because along the pathway of folding and association the polypeptide chain gains increasing stability [78, 79].

Several experimental approaches have been used to assign specific structural alterations to changes in stability: Selection of temperature-sensitive mutants; systematic variations of amino acid residues in the core or in the periphery of model proteins; cross-linking or joining of polypeptide chains; fragmentation of domain proteins or modifications of connecting peptides between domains; alterations of subunit interactions by mutagenesis or solvent perturbation [80]. As a result, protein stability has been found to be accomplished either by the covalent polypeptide backbone with or without disulfide-bridges, or by the above mentioned cumulative contributions of non-covalent interactions involved in the hierarchy of protein structure. The central issue in the adaptation of biomolecules to extreme conditions is the conservation of their functional state, which means a well-balanced compromise between stability and flexibility [81]. Basic mechanisms of molecular adaptation are changes in packing density, charge distribution, hydrophobic surface area, and in the ratio of polar or acidic: basic residues (acidic: basic residues). This indicates that the overall stability of a polypeptide chain must involve co-operativity, because the addition of stability increments per amino acid residue in the process of structure formation would not allow overcoming the thermal energy [82]. In the case of large multi domain or oligomeric proteins, kinetic partitioning, i.e. aggregation as a side reaction, competes with *in vitro* folding [83,84].

Determinants of Protein Folds

Secondary structure, the helices and sheets that are found in nearly every native protein structure, is stabilized primarily by hydrogen bonding between the amide and carbonyl groups of the main chain. The formation of such structure is an important element in the overall folding process, although it might not have as fundamental a role as the establishment of the overall chain topology [85]. At the base of the hierarchy are the regular secondary structures, e.g., α-helices and β-sheets, where consecutive residues adopt similar backbone conformations. Tertiary structure is then formed by packing secondary structural elements into one or several compact globular units called domains [86]. The latter is the average separation in the sequence between residues that are in contact with each other in the native structure. For proteins with more than about 100 residues, experiments generally reveal that one intermediate is significantly populated during the folding process. There has, however, been considerable discussion about the significance of such species: whether they assist the protein to find its correct structure or whether they are traps that inhibit the folding process [87]. Regardless of the outcome of this debate, the structural properties of intermediates provide important evidence about the folding of these larger proteins. In particular, they suggest that these proteins generally fold in modules, in other words, folding can take place independently in different segments or domains of the protein [88]. In such cases, interactions involving key residues are likely to establish the native-like fold within local regions or domains and also to ensure that the latter then interact appropriately to form the correct overall structure [89,90].

The rigid framework formed by secondary structures is the best-defined part of a protein structure. The spatial organization of secondary structural elements, or topology, has been the primary means by which protein structures and their commonalities are characterized and classified [91]. In addition to introducing order to the growing volume of structural data, the phenetic descriptions of protein structure also provide powerful clues to evolutionary relationships [92]. Empirical observations support the notion that structure is more robust than sequence [93]. Known protein folds differ markedly in the number of sequence families they can accommodate [94]. Although a majority of the folds have only one or two representatives in the current set of structurally characterized proteins, a small number of folds are associated with many unrelated families of sequences [86]. Native-like structure develops independently in the two distinct domains that make up the native fold: one domain is largely α-helical, and the other contains a substantial region of β-sheet [95,96].

In addition, structural regularities of proteins have long been recognized to be not only present at the whole protein (or domain) level, but also at the sub-structural level [97]. Secondary structural elements are observed to combine in specific geometric arrangements. The three basic super secondary structural motifs, α-hairpin, β-hairpin and βα-unit, are the simplest examples of such regularities [98]. Analysis of larger sub-structural motifs was expected to reveal more

about the topological preference of proteins [99]. With the amount of structural data currently available, our understanding of the topological preferences of sub-structural motifs has been greatly extended and some general principles can now be formulated. In particular, more than 50 % of protein domains consist of an open faced β-sheet flanked with helices or loops on either or both sides. Therefore, the topological biases at the sub-structural level directly influence the diversity of protein folds. Understanding these biases and the underlying physics helps in understanding why proteins prefer only a small fraction of the structural patterns [86].

What happens in the early stage of protein folding remains an interesting unsolved problem. Rapid kinetics measurements with cytochrome c using sub-millisecond continuous flow mixing devices suggest simultaneous formation of a compact collapsed state and secondary structure. These data seem to indicate that collapse formation is guided by specific short and long-range interactions (hetero-polymer collapse). A contrasting interpretation has also been proposed which suggests that the collapse formation is rapid, non-specific and a trivial solvent related compaction, which could as well be observed by a homo-polymer (homo-polymer collapse). The formation of secondary structure and compact state is not simultaneous in aqueous buffer. In aqueous buffer, formation of the compact state occurs through a two-state co-operative transition following hetero-polymer formalism while secondary structure formation takes place gradually. In contrast, in the presence of urea, a compaction of the protein radius occurs gradually over an extended range of salt concentration following homo-polymer formalism. The salt induced compaction and the formation of secondary structure take place simultaneously in the presence of urea [100].

Folds, Functions and Pathways

The native structure is an absolute requirement for protein function. Although knowing the fold alone usually does not give definite answers to all questions regarding function [101], rather the small number of basic protein folds provides a concise and powerful framework to organize the far larger number of biological functions needed by a living cell [102]. The major route of functional evolution is local mutation. Residues change as a protein evolves to satisfy modified functional constraints, while the basic biochemical mechanism and the overall three-dimensional fold remains unaltered. In most protein families, naturally occurring polymorphisms concentrate on residues that modulate the specificity of biological function [103]. Understanding how this is achieved and compiling a comprehensive mapping between protein folds and their related functions will be a major goal of structural biologists in the next few years. Although exploring the evolution of proteins and their functions in light of structural data is only just beginning, some fundamental relationships between folds and functions have already been revealed [104,105].

Elucidation of the structure and function of proteins, their interactions and the discovery of principles that provide unity to the enormous diversity of structural data, also have a deep impact on our understanding of the complex biochemical pathways. Structure mediates biological recognition, both within and between cells. The signals impinging on the cell surface are the inputs-the boundary conditions-that modulate a complex network of interactions within the cell. Much of the network connectivity, the links, is mediated by interactions between proteins. In this way, the behaviour of the cell can be seen to be qualitatively dependent on the local structure of one or more proteins. Selectively targeting these molecules based on knowledge of their structures would provide a way to control cell

behaviour, an approach that will reach its full power when a sufficient number of structures are available. Two distinct domains of native-like structure make the native fold: one is α-helical and another one is β-sheet. As one domain (the α-domain) folds faster than the other, a partially folded state with structure in the faster folding domain will result [106].

Moreover, current knowledge on the reaction whereby a protein acquires its native three-dimensional structure was obtained by and large through characterization of the folding mechanism of simple systems. Given the multiplicity of amino acid sequences and unique folds, it is not so easy, however, to draw general rules by comparing folding pathways of different proteins. In fact, quantitative comparison may be jeopardized not only because of the vast repertoire of sequences but also in view of a multiplicity of structures of the native and denatured states [107]. Fold recognition, assigning novel proteins to known structures, forms an important component of the overall protein structure discovery process. The available methods for protein fold recognition are limited by the low fold-coverage and / or low prediction accuracies. Mohammad & Nagarajaram [108] describe a new Support Vector Machine based method for protein fold prediction with high prediction accuracy and high fold-coverage. Since the new method gives rise to state of the art prediction performance and hence can be very useful for structural characterization of proteins discovered in various genomes.

Protein Unfolding Simulate Protein Folding

To accelerate unfolding by many orders of magnitude one has to mimic extreme denaturing conditions in one way or the other. Since stability is the main or even the only marker of the native fold the errors make it virtually impossible to find the native structure by a folding simulation starting with the unfolded chain or by any other molecular computations in the absence of additional artificial restraints [109]. In particular, simulations can help to identify or predict transition and intermediate states along the folding pathway, provide predictions of the rate of folding and in some cases, predict the final, folded structure. Although carried out under different (but always extreme) conditions, the simulations show a concurrent picture of early events in protein unfolding. The rupture of tight packing and liberation of side chains occur during the initial unfolding phase and then the water molecules penetrate inside the core and destroy, partly or completely, the secondary structure. The transition state for unfolding is found to be closer to the native than to the unfolded state. As a result of the early unfolding events, a protein usually adopts a collapsed, somewhat structured, molten globule-like state [110-112].

However, one should be careful not to overestimate the ability of molecular dynamics to simulate protein folding. Molecular dynamics can operate with nearly folded protein only, experimental studies point out in a very suggestive way that the most essential step of folding is the formation of a relatively small critical folding nucleus in the unfolded phase [113]. A discovery of a rather small folding nucleus [114] contradicts the earlier experimental evidence that the rate-limiting step of folding is near the native state [115]. A threat to the interpretation of such studies is that the folding pathway under strong folding conditions can differ from the unfolding pathway studied under strong unfolding conditions. According to the principle of detailed balance, a folding process must proceed via the same pathways as the reverse unfolding process when both of them are held at the same conditions, but the processes held under different conditions are not obliged to follow the same route. Molecular dynamics simulations aim to provide the most detailed description of the behaviour of 'realistic'

atomic models of proteins during their folding or unfolding processes. Molecular dynamics studies of protein folding have not yet come of age. It is hoped they will pay off in the future but, so far they have not given unexpected insights into folding or even unfolding mechanisms, but have rather presented detailed pictures of the processes already known 'in general'. Nevertheless, the all-atom simulations of protein unfolding can present interesting movies which awake our imagination and give rich and valuable information on the possible transition states and intermediates on the unfolding and possibly even the folding pathways [116,117].

The accurate characterization of the structure and dynamics of proteins in disordered states is a difficult problem at the frontier of structural biology whose solution promises to further our understanding of protein folding and intrinsically disordered proteins. Molecular dynamics simulations have added considerably to our understanding of folded proteins, but the accuracy with which the force fields used in such simulations can describe disordered proteins is unclear. Results of Subramani and Flouds demonstrate that the simulation successfully captures important aspects of both the local and global structure [118]. Molecular dynamics simulations can already be useful in describing disordered proteins. Direct calculation of certain NMR observables from the simulation provides new insight into the general relationship between structural features of disordered proteins and experimental NMR relaxation properties.

Future Prospective For Better Protein Folding

Molecular simulations are used to present the concept of establishing dynamic solution environments that mimic the molecular machinery employed for high efficiency protein folding *in vivo* to enhance the kinetic partitioning of the native conformation. In practice, the use of "SMART" polymers for protein folding in a decreasing temperature gradient mimics the capture-release mechanism of GroEL/GroES/ATP and promotes protein folding and inhibits protein aggregation. In normal circumstances the molecular chaperones and other 'housekeeping' mechanisms are remarkably efficient in ensuring that such potentially toxic species termed as prefibrillar aggregates are neutralized before they can do any damage [119]. This neutralization could result simply from the efficient targeting of misfolded proteins for degradation, but it seems that molecular chaperones are also able to alter the partitioning between harmful and harmless forms of aggregates (Figure 3) [5]. Oscillation of the oxidative/reductive potential of the solution by periodic loading of redox chemicals promotes the reshuffling of disulfide bridges, mimicking the action of protein disulfide isomerase and results in increased refolding yields. Realization of the simulated "oscillatory hydrophobic driving force" that mimics the quality control system in the ER may be of enormous practical value for protein folding at high concentrations [20].

Protein engineering techniques such as single point mutations, F-value analysis, C-value analysis and circular permutants are widely used to study protein folding pathways. These techniques also simultaneously provide possible ways to alter protein folding pathways *in vivo*, especially in the *E. coli* system. A denatured protein chain can find its well-ordered three-dimensional structure, the native state, in under a second, using only the information contained in the sequence. For researchers, however, the prediction of structures from sequences is a hard problem, so they are now recruiting all the help they can get, including idle computers and game consoles, game players and little hints from evolution [120].

Protein Fibrillation and Human Disease

Transition of soluble proteins and peptides into insoluble entities leads to the formation of fibrillar structure. These fibrils are characterized by a predominant cross β-sheet conformation where the β-strands run parallel to the fibril axis [121]. The fibrillar structure if formed intracellularly is termed as inclusion bodies and formation in the extracellular medium leads to amyloid structure (Figure 4) [122]. A range of human diseases are associated with fibrillation of proteins. These include both neurodegenerative and non- neurodegenerative dysfunctions. Among the first group of diseases mentioned we can include the fibrillation of Aβ peptide, Parkinson's involving the aggregation and fibrillation of α-synuclein. As far as the non-neurodegenerative diseases are concerned there is a list of disorders which include cataract, cystic fibrosis, chronic pancreatitis etc [123].
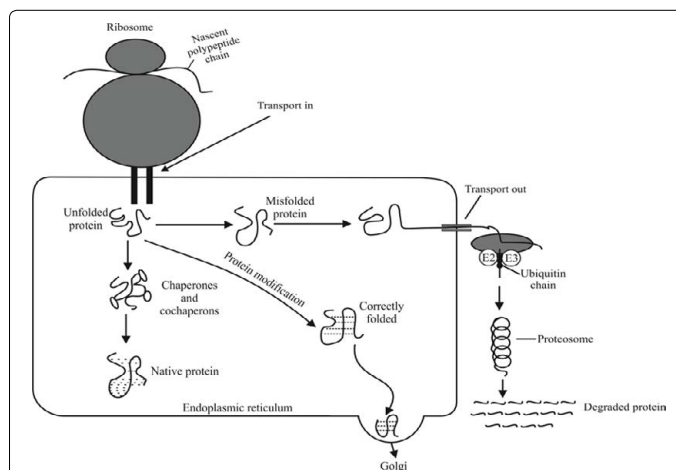

Figure 3: Protein folding regulation in the endoplasmic reticulum. Newly synthesized proteins are translocated into the endoplasmic reticulum, where they fold into their three-dimensional structure with the help of chaperones. Correctly folded proteins transported to golgi complex. However, incorrectly folded proteins are ubiquitinated and than degraded into cytoplasm by proteosomes (E2, ubiquitin-conjugating enzyme; E3, ubiquitin ligase).
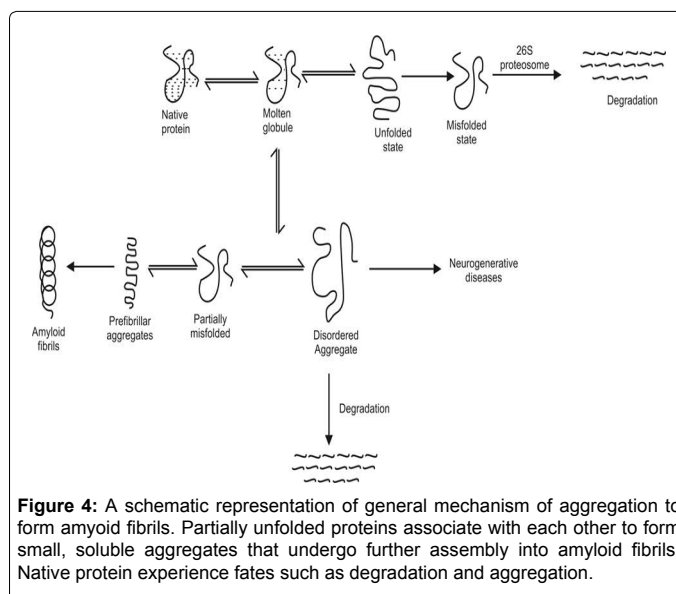

Figure 4: A schematic representation of general mechanism of aggregation to form amyoid fibrils. Partially unfolded proteins associate with each other to form small, soluble aggregates that undergo further assembly into amyloid fibrils. Native protein experience fates such as degradation and aggregation.

Conclusion

It is therefore essential that we use our advanced understanding of protein misfolding and aggregation to find effective strategies for combating these increasingly common and highly debilitating diseases. Fortunately, there is now real evidence to suggest that modern science will rise successfully to this tremendous challenge. Given the importance of topology in determining the protein-folding mechanism, complete knowledge of the core folding units will facilitate the development of more effective methods for protein structure prediction. An understanding of folding is important for the analysis of many events involved in cellular regulation, the design of proteins with novel functions, and the utilization of sequence information from the various genome projects and the development of novel therapeutic strategies for treating or preventing debilitating human diseases that are associated with the failure of proteins to fold correctly. The fundamental principles of protein folding have practical applications in the exploitation of advances in genomic research. Detailed mechanistic understanding of protein aggregation and amyloid formation *in vitro* and *in vivo* presents several challenges that remain to be addressed. Several fundamental questions about the molecular and structural determinants of amyloid formation and the mechanisms of amyloid-induced toxicity remain unanswered.

Acknowledgement

Conflicts of interest

The authors declare that there is no conflict of interest.

References

1. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, et al. (2005) Evolutionary information for specifying a protein fold. Nature 437: 512-518.

2. Pappu RV, Nussinov R (2009) Protein folding: lessons learned and new frontiers. Preface. Phys Biol 6: 010301.

3. Radford SE, Dobson CM (1999) From computer simulations to human disease: emerging themes in protein folding. Cell 97: 291-298.

4. Li W, Zhang J, Wang J, Wang W (2008) Metal-coupled folding of Cys2His2 zinc-finger. J Am Chem Soc 130: 892-900.

5. Nicastro I, Dillin A (2013) Proteostasis and the aging pathways. In protein quality control in neurodegenerative diseases, Springer Berlin Heidelberg 37-52.

6. Ben-Naim A (2011) Some aspects of the protein folding problem examined in one-dimensional systems. J Chem Phys 135: 085104.

7. Crevenna AH, Naredi-Rainer N, Lamb DC, Wedlich-Söldner R, Dzubiella J (2012) Effects of Hofmeister ions on the Î±-helical structure of proteins. Biophys J 102: 907-915.

8. Dobson CM, Sali A, Karplus M (1998) Protein Folding: a perspective from theory and experiment. Angew Chem Int Ed 37: 868-893.

9. Kamerlin SC, Vicatos S, Dryga A, Warshel A (2011) Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. Annu Rev Phys Chem 62: 41-64.

10. Marracino P, Apollonio F, Liberti M, d'Inzeo G, Amadei A (2013) Effect of high exogenous electric pulses on protein conformation: myoglobin as a case study. J Phys Chem B 117: 2273-2279.

11. Dobson CM, Evans PA, Radford SE (1994) Understanding how proteins fold: the lysozyme story so far. Trends Biochem Sci 19: 31-37.

12. Bryngelson JD, Wolynes PG (1989) Intermediates and barriers crossing in a random energy model (application to protein folding). J Phys Chem 93: 6902-6915.

13. Yang LQ, Ji XL, Liu SQ (2013) The free energy landscape of protein folding and dynamics: a global view. J Biomol Struct Dyn 31: 982-992.

14. de Groot NS, Espargaró A, Morell M, Ventura S (2008) Studies on bacterial inclusion bodies. Future Microbiol 3: 423-435.

15. Bae JY, Koo BK, Ryu HB, Song JA, Nguyen MT, et al. (2013) Cu/Zn incorporation during purification of soluble human EC-SOD from E. coli stabilizes proper disulfide bond formation. Appl Biochem Biotechnol 169: 1633-1647.

16. Predeus AV, Gul S, Gopal SM, Feig M (2012) Conformational sampling of peptides in the presence of protein crowders from AA/CG-multiscale simulations. J Phys Chem B 116: 8610-8620.

17. Miranker AD, Dobson CM (1996) Collapse and cooperativity in protein folding. Curr Opin Struct Biol 6: 31-42.

18. Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181: 223-230.

19. Daggett V, Fersht A (2003) The present view of the mechanism of protein folding. Nat Rev Mol Cell Biol 4: 497-502.

20. Lu D, Liu Z (2010) Studies of protein folding pathways. Annu Rep Prog Chem Sect C: Phys Chem 106: 259-273.

21. Radley TL, Markowska AI, Bettinger BT, Ha JH, Loh SN (2003) Allosteric switching by mutually exclusive folding of protein domains. J Mol Biol 332: 529-536.

22. Gething MJ, Sambrook J (1992) Protein folding in the cell. Nature 355: 33-45.

23. Mallam AL, Jackson SE (2011) Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. Nat Chem Biol 8: 147-153.

24. Bajaj M, Blundell T (1984) Evolution and the tertiary structure of proteins. Annu Rev Biophys Bioeng 13: 453-492.

25. Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5: 823-826.

26. Perutz MF, Kendrew JC, Watson HC (1965) Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence. J Mol Biol 13: 669-678.

27. Zastrow ML, Pecoraro VL (2013) Designing functional metalloproteins: from structural to catalytic metal sites. Coord Chem Rev 257: 2565-2588.

28. McCammon JA, Harvey SC (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, London.

29. Smith JL, Hendrickson WA, Honzatko RB, Sheriff S (1986) Structural heterogeneity in protein crystals. Biochemistry 25: 5018-5027.

30. Wagner G (1983) Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes. Q Rev Biophys 16: 1-57.

31. Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E, Razgulyaev OI (1990) Evidence for a molten globule state as a general intermediate in protein folding. FEBS Lett 262: 20-24.

32. Juneja J, Udgaonkar JB (2003) NMR studies of protein folding. Curr Sci 84: 157-172.

33. Devaraneni PK, Mishra N, Bhat R (2012) Polyol osmolytes stabilize native-like cooperative intermediate state of yeast hexokinase A at low pH. Biochimie 94: 947-952.

34. Denny RA, Gavrin LK, Saiah E (2013) Recent developments in targeting protein misfolding diseases. Bioorg Med Chem Lett 23: 1935-1944.

35. Figueiredo AM, Moore GR, Whittaker SB (2012) Understanding how small helical proteins fold: conformational dynamics of Im proteins relevant to their folding landscapes. Biochem Soc Trans 40: 424-428.

36. Privalov PL (1989) Thermodynamic problems of protein structure. Annu Rev Biophys Biophys Chem 18: 47-69.

37. Rose GD, Fleming PJ, Banavar JR, Maritan A (2006) A backbone-based theory of protein folding. Proc Natl Acad Sci U S A 103: 16623-16633.

38. Chang JY (2009) Structural heterogeneity of 6 M GdmCl-denatured proteins: implications for the mechanism of protein folding. Biochemistry 48: 9340-9346.

39. Tran HT, Pappu RV (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. Biophys J 91: 1868-1886.

40. Dar TA, Schaeffer RD, Daggett V, Bowler BE (2011) Manifestations of native topology in the denatured state ensemble of Rhodopseudomonas palustris cytochrome c'. Biochemistry 50: 1029-1041.

41. Zambelli B, Cremades N, Neyroz P, Turano P, Uversky VN, et al. (2012) Insights in the (un)structural organization of Bacillus pasteurii UreG, an intrinsically disordered GTPase enzyme. Mol Biosyst 8: 220-228.

42. Kuwajima K (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. Proteins 6: 87-103.

43. Bhattacharya M, Mukhopadhyay S (2012) Structural and dynamical insights into the molten-globule form of ovalbumin. J Phys Chem B 116: 520-531.

44. Creighton TE (1997) Protein folding coupled to disulphide bond formation. Biol Chem 378: 731-744.

45. Mezzenga R, Fischer P (2013) The self-assembly, aggregation and phase transitions of food protein systems in one, two and three dimensions. Rep Prog Phys 76: 046601.

46. Wu Y, Kondrashkina E, Kayatekin C, Matthews CR, Bilsel O (2008) Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. Proc Natl Acad Sci U S A 105: 13367-13372.

47. Ishii T, Murayama Y, Katano A, Maki K, Kuwajima K, et al. (2008) Probing force-induced unfolding intermediates of a single staphylococcal nuclease molecule and the effect of ligand binding. Biochem Biophys Res Commun 375: 586-591.

48. Gsponer J, Hopearuoho H, Whittaker SB, Spence GR, Moore GR, et al. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. Proc Natl Acad Sci U S A 103: 99-104.

49. Freddolino PL, Schulten K (2009) Common structural transitions in explicit-solvent simulations of villin headpiece folding. Biophys J 97: 2338-2347.

50. Elms PJ, Chodera JD, Bustamante C, Marqusee S (2012) The molten globule state is unusually deformable under mechanical force. Proc Natl Acad Sci U S A 109: 3796-3801.

51. Moosavi-Movahedi AA, Chamani J, Goto Y, Hakimelahi GH (2003) Formation of the molten globule-like state of cytochrome c induced by n-alkyl sulfates at low concentrations. J Biochem 133: 93-102.

52. Kelley EG, Albert JN, Sullivan MO, Epps TH 3rd (2013) Stimuli-responsive copolymer solution and surface assemblies for biomedical applications. Chem Soc Rev 42: 7057-7071.

53. Dill KA, Ozkan SB, Shell MS, Weikl TR (2008) The protein folding problem. Annu Rev Biophys 37: 289-316.

54. Anfinsen CB, Scheraga HA (1975) Experimental and theoretical aspects of protein folding. Adv Protein Chem 29: 205-300.

55. Yang JS, Chen WW, Skolnick J, Shakhnovich EI (2007) All-atom ab initio folding of a diverse set of proteins. Structure 15: 53-63.

56. Chen J, Stites WE (2001) Packing is a key selection factor in the evolution of protein hydrophobic cores. Biochemistry 40: 15280-15289.

57. Fiedler S, Broecker J, Keller S (2010) Protein folding in membranes. Cell Mol Life Sci 67: 1779-1798.

58. Wolfenden R (2007) Experimental measures of amino acid hydrophobicity and the topology of transmembrane and globular proteins. J Gen Physiol 129: 357-362.

59. Kim DE, Gu H, Baker D (1998) The sequences of small proteins are not extensively optimized for rapid folding by natural selection. Proc Natl Acad Sci U S A 95: 4982-4986.

60. Wurth C, Kim W, Hecht MH (2006) Combinatorial approaches to probe the sequence determinants of protein aggregation and amyloidogenicity. Protein Pept Lett 13: 279-286.

61. Privalov PL, Tiktopulo EI, Venyaminov SYu, Griko YuV, Makhatadze GI, et al. (1989) Heat capacity and conformation of proteins in the denatured state. J Mol Biol 205: 737-750.

62. Mikulecky PJ, Feig AL (2004) Heat capacity changes in RNA folding: application of perturbation theory to hammerhead ribozyme cold denaturation. Nucleic Acids Res 32: 3967-3976.

63. Creighton TE (1983) Proteins: Structures and Molecular Properties, WH Freeman, New York.

64. Muller N (1990) Search for a realistic view of hydrophobic effects. Acc Chem Res 23: 23-28.

65. Bello J (1977) Stability of protein conformation: internal packing and enthalpy of fusion of model compounds. J Theor Biol 68: 139-142.

66. Anderson DE, Becktel WJ, Dahlquist FW (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemistry 29: 2403-2408.

67. Luo Y, Baldwin RL (1998) Trifluoroethanol stabilizes the pH 4 folding intermediate of sperm whale apomyoglobin. J Mol Biol 279: 49-57.

68. Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295: 1852-1858.

69. Vabulas RM, Raychaudhuri S, Hayer-Hartl M, Hartl FU (2010) Protein folding in the cytoplasm and the heat shock response. Cold Spring Harb Perspect Biol 2: a004390.

70. Schwarz FP (1988) Interaction of cytidine 3'-monophosphate and uridine 3'-monophosphate with ribonuclease a at the denaturation temperature. Biochemistry 27: 8429-8436.

71. Liu W, Tsou CL (1987) Kinetics of reactivation during refolding of guanidine-denatured pancreatic ribonuclease A. Biochim Biophys Acta 916: 465-473.

72. Fu Y, Kasinath V, Moorman VR, Nucci NV, Hilser VJ, et al. (2012) Coupled motion in proteins revealed by pressure perturbation. J Am Chem Soc 134: 8543-8550.

73. Cheung MS, García AE, Onuchic JN (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc Natl Acad Sci U S A 99: 685-690.

74. Li J, Hoop CL, Kodali R, Sivanandam VN, van der Wel PC (2011) Amyloid-like fibrils from a domain-swapping protein feature a parallel, in-register conformation without native-like interactions. J Biol Chem 286: 28988-28995.

75. Dill KA (1990) Dominant forces in protein folding. Biochemistry 29: 7133-7155.

76. Pfeil W (1998) Protein stability and folding Supplement 1: a collection of thermodynamic data (1st edn). Springer-Verlag Berlin Heidelberg, NewYork.

77. Makyla K, Müller C, Lörcher S, Winkler T, Nussbaumer MG, Eder M, Bruns N (2013) Fluorescent protein senses and reports mechanical damage in glass-Fiber-Reinforced polymer composites. Adv Mater, 25: 2701-2706.

78. Jaenicke R (1999) Stability and folding of domain proteins. Prog Biophys Mol Biol 71: 155-241.

79. Zhou H, Zhou Y (2004) Quantifying the effect of burial of amino acid residues on protein stability. Proteins 54: 315-322.

80. Matthews BW (1996) Structural and genetic analysis of the folding and function of T4 lysozyme. FASEB J 10: 35-41.

81. Tsou CL (1988) Folding of the nascent peptide chain into a biologically active protein. Biochemistry 27: 1809-1812.

82. Jaenicke R, Lilie H (2000) Folding and association of oligomeric and multimeric proteins. Adv Protein Chem 53: 329-401.

83. Jaenicke R (1996) Protein folding and association: in vitro studies for self-organization and targeting in the cell. Curr Top Cell Regul 34: 209-314.

84. Sabate R, de Groot NS, Ventura S (2010) Protein folding and aggregation in bacteria. Cell Mol Life Sci 67: 2695-2715.

85. Krobath H, Faísca PF (2013) Interplay between native topology and non-native interactions in the folding of tethered proteins. Phys Biol 10: 016002.

86. Zhang C, DeLisi C (2001) Protein folds: molecular systematics in three dimensions. Cell Mol Life Sci 58: 72-79.

87. Khan F, Chuang JI, Gianni S, Fersht AR (2003) The kinetic pathway of folding of barnase. J Mol Biol 333: 169-186.

88. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 25: 331-339.

89. Vendruscolo M, Paci E, Karplus M, Dobson CM (2003) Structures and relative free energies of partially folded states of proteins. Proc Natl Acad Sci U S A 100: 14817-14821.

90. Sree BR, Joy JM (2011) A review on molecular and chaperone overload. Int J Pharm Dev Technol 1: 58-65.

91. Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv Protein Chem 34: 167-339.

92. Holm L (1998) Unification of protein families. Curr Opin Struct Biol 8: 372-379.

93. Wood TC, Pearson WR (1999) Evolution of protein sequences and structures. J Mol Biol 291: 977-995.

94. Brenner SE, Chothia C, Hubbard TJ (1997) Population statistics of protein structures: lessons from structural classifications. Curr Opin Struct Biol 7: 369-376.

95. Miranker A, Radford SE, Karplus M, Dobson CM (1991) Demonstration by NMR of folding domains in lysozyme. Nature 349: 633-636.

96. Dobson CM (2001) The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci 356: 133-145.

97. Branden C, Tooze J (1999) Introduction to Protein Structure. Garland, New York.

98. Sibanda BL, Thornton JM (1985) Beta-hairpin families in globular proteins. Nature 316: 170-174.

99. Efimov AV (1994) Common structural motifs in small proteins and domains. FEBS Lett 355: 213-219.

100. Haldar S, Chattopadhyay K (2012) Interconnection of salt-induced hydrophobic compaction and secondary structure formation depends on solution conditions: revisiting early events of protein folding at single molecule resolution. J Biol Chem 287: 11546-11555.

101. Martin AC, Orengo CA, Hutchinson EG, Jones S, Karmirantzou M, et al. (1998) Protein folds and functions. Structure 6: 875-884.

102. Thornton JM, Orengo CA, Todd AE, Pearl FM (1999) Protein folds, functions and evolution. J Mol Biol 293: 333-342.

103. Parham P, Lomen CE, Lawlor DA, Ways JP, Holmes N, et al. (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. Proc Natl Acad Sci U S A 85: 4005-4009.

104. Orengo CA, Todd AE, Thornton JM (1999) From protein structure to function. Curr Opin Struct Biol 9: 374-382.

105. Atkinson HJ, Babbitt PC (2009) An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. PLoS Comput Biol 5: e1000541.

106. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37: 215-246.

107. Brunori M, Gianni S, Giri R, Morrone A, Travaglini-Allocatelli C (2012) Morphogenesis of a protein: folding pathways and the energy landscape. Biochem Soc Trans 40: 429-432.

108. Mohammad TA, Nagarajaram HA (2011) A hierarchical approach to protein fold prediction. J Integr Bioinform 8: 185.

109. De Sancho D, Mittal J, Best RB (2013) Folding kinetics and unfolded state dynamics of the GB1 hairpin from molecular simulation. J Chem Theory Comput 9: 1743-1753.

110. Finkelstein AV (1997) Can protein unfolding simulate protein folding? Protein Eng 10: 843-845.

111. Maiti NR, Surewicz WK (2001) The role of disulfide bridge in the folding and stability of the recombinant human prion protein. J Biol Chem 276: 2427-2431.

112. Hill JJ, Shalaev EY, Zografi G (2005) Thermodynamic and dynamic factors involved in the stability of native protein structure in amorphous solids in relation to levels of hydration. J Pharm Sci 94: 1636-1667.

113. Fersht AR (1997) Nucleation mechanisms in protein folding. Curr Opin Struct Biol 7: 3-9.

114. Fersht AR (1995) Characterizing transition states in protein folding: an essential step in the puzzle. Curr Opin Struct Biol 5: 79-84.

115. Creighton TE (1988) Toward a better understanding of protein folding pathways. Proc Natl Acad Sci U S A 85: 5082-5086.

116. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. Curr Opin Struct Biol 14: 76-88.

117. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. Science 334: 517-520.

118. Subramani A, Floudas CA (2012) Structure prediction of loops with fixed and flexible stems. J Phys Chem B 116: 6670-6682.

119. Sherman MY, Goldberg AL (2001) Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. Neuron 29: 15-32.

120. Gross M (2012) Folding research recruits unconventional help. Curr Biol 22: R35-38.

121. Ramshini H, Parrini C, Relini A, Zampagni M, Mannini B, et al. (2011) Large proteins have a great tendency to aggregate but a low propensity to form amyloid fibrils. PLoS One 6: e16075.

122. Thomas SI, Balaji PV (2007) Protein aggregation: a perspective from amyloid and inclusion-body formation. Curr sci 92: 758-767.

123. Naeem A, Fazili NA (2011) Defective protein folding and aggregation as the basis of neurodegenerative diseases: the darker aspect of proteins. Cell Biochem Biophys 61: 237-250.