

# Theoretical Approach to the Common Events in Every Living Cell - Protein Folding and Protein Misfolding

V. K. Vaibhav<sup>1,\*</sup>, S. L. Sachin<sup>2</sup>

<sup>1</sup>Department of Pharmaceutical Biotechnology, Svb's College of Pharmacy, Kalyan-Shill Road, Dombivli (E)-421204, Maharashtra, India

<sup>2</sup>Department of Pharmaceutical Chemistry, Ideal College of Pharmacy and Research, Kalyan-(W)-421206, Maharashtra, India

\*E-mail address: vaibhav16528@gmail.com

## ABSTRACT

Folding and unfolding are crucial ways of regulating biological activity and targeting proteins to different cellular locations. 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. Protein misfolding is a common event in living cells. Molecular chaperones not only assist protein folding; they also facilitate the degradation of misfolded polypeptides. Protein folding is governed solely by the protein itself, scientists discovered that some proteins have helped in the process called chaperones. When the intracellular degradative capacity is exceeded, juxtannuclear aggresomes are formed to sequester misfolded proteins. Misfolding of newly formed proteins not only results in a loss of physiological function of the protein but also may lead to the intra- or extra- cellular accumulation of that protein. A number of diseases have been shown to be characterised by the accumulation of misfolded proteins, notable example being Alzheimer's disease.

## Keywords:

Protein Folding; Protein Misfolding; Molecular Chaperones; Diseases

## 1. INTRODUCTION

All computational models that predict something have certain underlying assumptions that constitute the physical basis for the model. In protein structure prediction, there are two physical/biological processes that can be modeled: the process of evolution, or the process of folding. We may give these two paradigms names, Darwin and Boltzmann, after the scientists who defined the fundamental principles of evolutionary biology and statistical thermodynamics, respectively. Most of the work in protein structure prediction is Darwin-based, using the well-known premise that sequences that have a common ancestor have similar folds, and they strive to extrapolate this principle to increasingly distant sequence relationships. Methods that use multiple sequence alignment, structural alignment, or "threading potentials" are implicitly searching for a common ancestor. Despite the oft-used "energy-like" scoring functions, these methods do not address the physical process of folding. Evolution happens on the time scale of millions of years, folding on the time scale of fractions of a second. Protein structure prediction of the Boltzmann

kind is perceived to be a very difficult problem. Many have tried their hand at it over the last thirty years, and an equal number have failed to improve upon Darwin-based methods. The problem of predicting folding pathways may be perceived to be even harder, since it should depend on first solving the protein folding problem. But this is not true, as we shall see. Prediction of the protein folding pathway may be evaluated by looking at the success in predicting sub-segments or substructures of proteins. If the computational model has the right underlying assumptions about what comes first in the pathway, and what comes next, and so on, then blind predictions, such as those done as part of CASP, the Critical Assessment of Protein Structure Prediction biannual worldwide experiment ; may validate that model. And the pathway model that eventually arises from this process will tell us more than just final answer.[1]

## 2. PROTEIN FOLDING PATHWAY HISTORY

The early work of Levinthal and Anfinsen established that a protein chain folds spontaneously and reproducibly to a unique three dimensional structure when placed in aqueous solution. Levinthal proved beyond the shadow of a doubt that the folding process cannot occur by random diffusion. Anfinsen proposed that proteins must form intermediate structures in a time-ordered sequence of events, or "pathway". The nature of the pathways, specifically whether they are restricted to partially native states or whether they might Modeling Protein Folding Pathways 3 include non-specific interactions, such as an early collapse driven by the hydrophobic effect, was left unanswered. Over the years, the theoretical models for folding have converged somewhat, in part due to a better understanding of the structure of the so-called "unfolded state" and to a more detailed description of kinetic and equilibrium folding intermediates . An image of the transition state of folding can now be mapped out by point mutations, or "phi-value analysis". The "folding funnel" model has reconciled hydrophobic collapse with the alternative nucleation-condensation model by envisioning a distorted, funicular energy landscape and a "minimally frustrated" pathway through this landscape. The view remains of a channeled, counter-entropic search for the hole in the funnel as the predominant barrier to folding. Simulations using various simplified representations of the protein chain, including lattice models, have clarified the basic nature of folding pathways. The topology of the fold plays a dominant role in defining the critical positions that affect the folding rate. Models that represent the chain in atomistic detail show that minimally frustrated, low-energy pathways may involve the propagation of structure along the chain like a zipper. All-atom, explicit solvent molecular dynamics simulations have reproduced the experimentally determined conformations for short peptides. This large body of work is still inconclusive, but clearly folding is best represented by an ensemble rather than a single pathway. [1]

## 3. PROTEIN FOLDING

Protein folding is the process by which a string of amino acids (the chemical building blocks of protein) interacts with itself to form a stable three-dimensional structure during production of the protein within the cell. The process is roughly analogous to the ways in which a length of wire may be twisted onto or against itself to form various functional entities, for example a spring, a paperclip or a coathanger. Folding occurs very rapidly, probably within milliseconds of production of the string of amino acids, and results in 3-D conformations which usually are quite stable, with specific biological functions. [12] The folding of proteins thus facilitates the

production of discrete functional entities, including enzymes and structural proteins, which allow the various processes associated with life to occur. Importantly, folding not only allows the production of structures which can perform particular functions in the cellular milieu, but also it prevents inappropriate interactions between proteins, in that folding hides elements of the amino acid sequence which if exposed would react non-specifically with other proteins. Restriction of interactions to those which are necessary and desirable for life is crucial in the intracellular environment where many thousands of proteins are present and required to perform precisely specified functions. Evolutionary pressure thus has favoured those proteins which fold in such a way that appropriate reactive elements are exposed and unwanted reactivities are hidden. [1, 2].

#### 4. HOW DO PROTEINS FOLD INSIDE CELLS?

As well as enzymes that isomerize covalent bonds in protein chains, cells contain a variety of molecular chaperones that control and assist the folding process. Previous work suggests that two types of chaperone act sequentially on newly synthesized polypeptides in both the cytoplasm of prokaryotic cells and in the cytosol and mitochondria of eukaryotic cells. Small chaperones (100 kDa), such as hsp70 (DnaK) and hsp40 (DnaJ), bind to hydrophobic regions on nascent chains to prevent aggregation and premature folding as elongation continues, while large chaperones (800 kDa), such as GroEL, bind complete, partially folded chains individually in a central cage, where folding proceeds further until the danger of aggregation with similar chains has passed. [2]

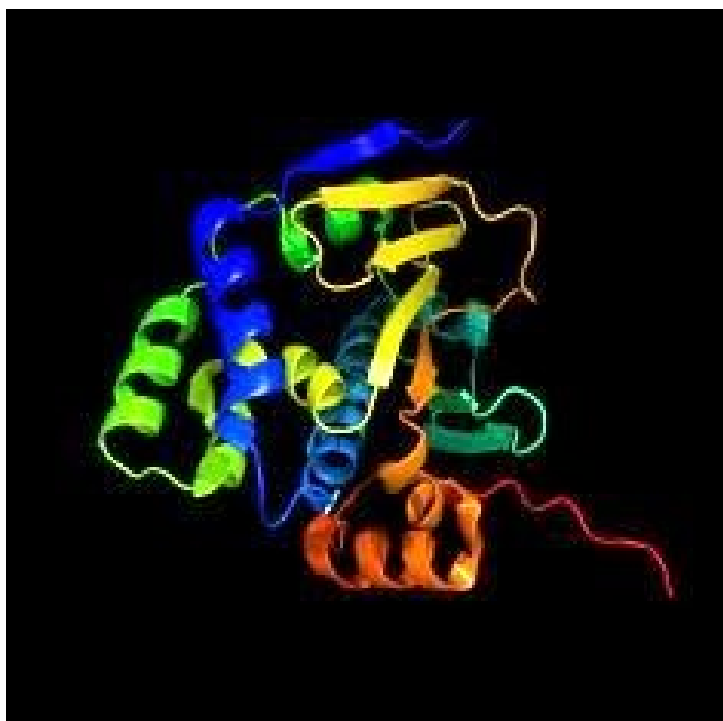


Fig. 1. Folded Protein

## 5. WHAT SHAPE WILL A PROTEIN FOLD INTO?

Even though proteins are just a long chain of amino acids, they don't like to stay stretched out in a straight line. The protein folds up to make a compact blob, but as it does, it keeps some amino acids near the center of the blob, and others outside; and it keeps some pairs of amino acids close together and others far apart. Every kind of protein folds up into a very specific shape - the same shape every time. Most proteins do this all by themselves, although some need extra help to fold into the right shape. The unique shape of a particular protein is the most stable state it can adopt. Picture a ball in a funnel - the ball will always roll down to the bottom of the funnel because that is the most stable state [3].

## 6. WHY IS SHAPE IMPORTANT?

This structure specifies the function of the protein. For example, a protein that breaks down glucose so the cell can use the energy stored in the sugar will have a shape that recognizes the glucose and binds to it (like a lock and key) and chemically reactive amino acids that will react with the glucose and break it down to release the energy[1,3].

## 7. THE PROTEIN FOLDING PROCESS

The folding pathway of a large polypeptide chain is very complicated, and not all the principles that guide the process have been worked out. However, many plausible models have attempted to describe protein folding. One model views folding as a hierarchical process where local secondary structures form first. Under this model,  $\alpha$ -helices and  $\beta$ -sheets form first, with longer range interactions between helices and sheets forming super-secondary structures later. This process continues until the entire polypeptide folds. An alternative model describes folding as a spontaneous collapse of the polypeptide into a compact state. This collapsed state is known as a molten globule. It may be that the actual folding process of proteins incorporates features of both models. Instead of following a single pathway, a population of peptide molecules may take a variety of routes. Thermodynamically, the folding process can be viewed as a kind of free-energy funnel, where the unfolded states are characterized by a high degree of conformational entropy and relatively high free energy. In a trivialized definition, entropy is a measure of chaos, a measure of all different conformational states that the protein can be in. Obviously, there is more chaos in the protein in its unfolded state.

On the other hand high free energy is a measure of unstableness, which is higher in a protein's unfolded state. Therefore, as folding proceeds, the narrowing of the funnel represents a decrease in the number of conformational states present. Local minima along the sides of the free energy funnel represent transition states that are semistable and can briefly slow the protein states since it takes some time for the protein to jump out of the local minima. At the bottom of the funnel, also known as the global minimum, an ensemble of folding intermediates are reduced to a single conformation. It is important to realize that although we often describe the free energy funnel as having one global minimum - that is, one native conformation - a protein can have a small set of native conformations, each one important for its biological function(s). [3]

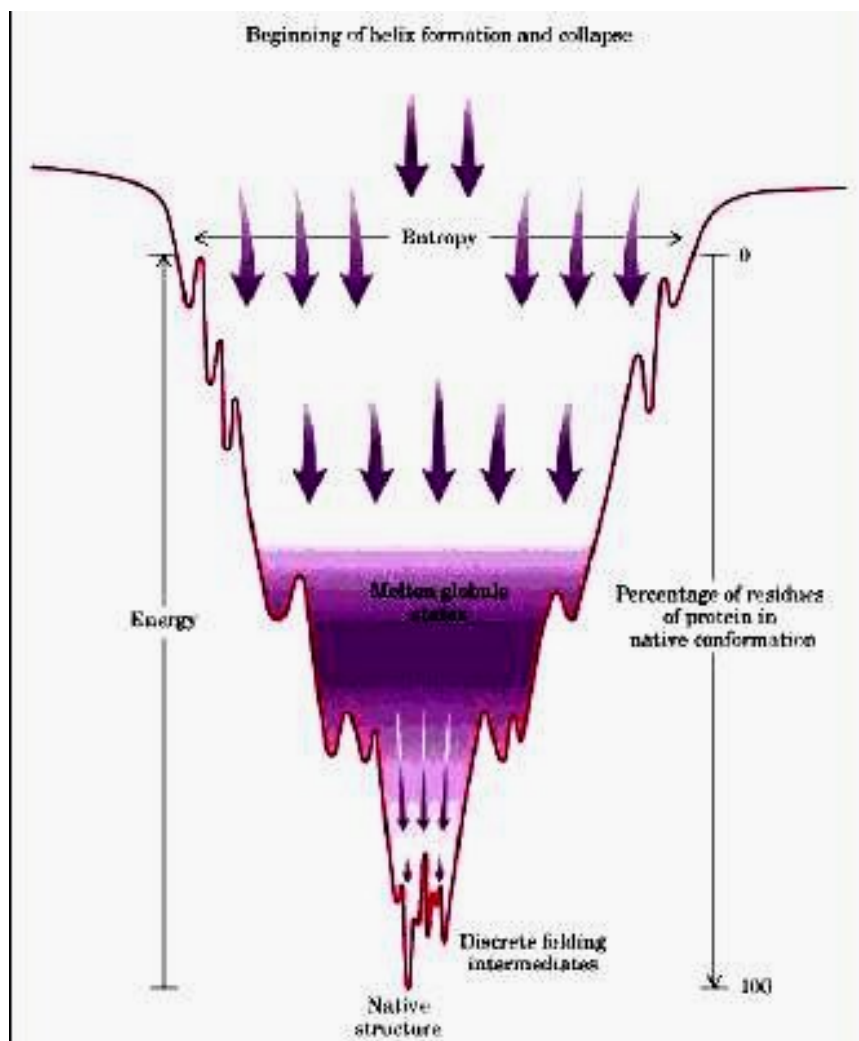


Fig. 2. Free Energy Funnel

## 8. THE 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. Perhaps the most dramatic evidence for such a conclusion is the observation of a remarkable correlation between the experimental folding rates of a wide range of small proteins and the complexity of their folds, measured by the contact order. [12] The latter is the average separation in the sequence between residues that are in contact with each other in the native structure. The existence of such a correlation can be rationalized by the argument that a stochastic search process will be more time consuming if the residues that form the nucleus are further away from each other in the sequence. This evidence strongly supports the conclusion that there are

relatively simple underlying principles by which the sequence of a protein encodes its structure. Not only will the establishment of such principles reveal in more depth how proteins are able to fold, but it should advance significantly our ability to predict protein folds directly from their sequences and to design sequences that encode novel folds.

For proteins with more than about 100 residues, experiments generally reveal that one (or more) 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. 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 largely independently in different segments or domains of the protein. [4]

## 9. PROTEIN FOLDING AND MISFOLDING IN THE CELL

In a cell, proteins are synthesized on ribosomes from the genetic information encoded in the cellular DNA. Folding *in vivo* is in some cases co-translational; that is, it is initiated before the completion of protein synthesis, whereas the nascent chain is still attached to the ribosome. Other proteins, however, undergo the major part of their folding in the cytoplasm after release from the ribosome, whereas yet others fold in specific compartments, such as mitochondria or the endoplasmic reticulum (ER), after trafficking and translocation through membranes.

Many details of the folding process depend on the particular environment in which folding takes place, although the fundamental principles of folding, discussed above, are undoubtedly universal. But because incompletely folded proteins must inevitably expose to the solvent at least some regions of structure that are buried in the native state, they are prone to inappropriate interaction with other molecules within the crowded environment of a cell. Living systems have therefore evolved a range of strategies to prevent such behavior of particular importance in this context are the many molecular chaperones that are present in all types of cells and cellular compartments. Some chaperones interact with nascent chains as they emerge from the ribosome, whereas others are involved in guiding later stages of the folding process.

Molecular chaperones often work in tandem to ensure that the various stages in the folding of such systems are all completed efficiently. Many of the details of the functions of molecular chaperones have been determined from studies of their effects on folding *in vitro*. The best characterized of the chaperones studied in this manner is the bacterial complex involving GroEL, a member of the family of 'chaperonins', and its 'co-chaperone' GroES. Many aspects of the sophisticated mechanism through which this coupled system functions are now well understood.

Of particular interest are that GroEL, and other members of this class of molecular chaperone, contains a cavity in which incompletely folded polypeptide chains can enter and undergo the final steps in the formation of their native structures while sequestered and protected from the outside world. [5, 6]

## 10. MOLECULAR CHAPERONES

Protein folding is governed solely by the protein itself, some proteins have helped in the process. This help consists of proteins called chaperones (or chaperonins) that are associated with the target protein during part of its folding process. However, once folding is complete (or even before) the chaperone will leave its current protein molecule and go on to support the folding of another. Proper folding of some proteins appears to call for not just one chaperone, but several. Especially clear evidence for such multi-step chaperoning is provided by test-tube experiments on a protein known as rhodanese. Proper folding of this protein, the experiments show, requires five different chaperone-type proteins acting at two distinct steps in the operation. Early in the folding process, rhodanese binds to a chaperone known as DnaK; the complex that binds a further chaperone :DnaJ. Somewhat later, a protein known as GrpE catalyzes transfer of the partially folded rhodanese to another chaperone, GroEL, and its partner, GroES. These latter two proteins then see rhodanese all the way through to its properly folded state. Several lines of evidence suggest that Chaperones. Primary function may be to prevent aggregation. For example, a chaperone found in the power plant. Organelles of mammalian cells (but otherwise similar to GroEL) has been shown to consist of 14 protein chains arranged as two doughnuts stacked on top of each other (figure 2). The chaperoned protein sits inside the two doughnut holes, safely sequestered from other molecules with which it might aggregate.

A role for chaperones in preventing aggregation is also suggested by what happens to mammalian proteins produced in bacteria. Although bacteria have chaperones, they are not the same as those in mammals. It is thus easy to imagine that they may be relatively ineffective toward mammalian proteins, and that this results in the aggregation so often seen. Indeed, there has been one case in which bacteria engineered to overproduce their own chaperones successfully produced a mammalian protein that otherwise irretrievably aggregated. Unfortunately, this approach has failed in other cases. And no one has yet reported introduction of mammalian chaperones into bacteria to help produce soluble mammalian proteins. [2, 9, 10].

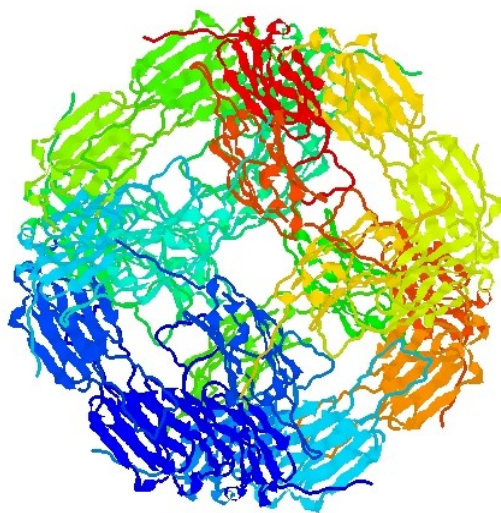


Fig. 3. Molecular Chaperones

## 11. PROTEIN MISFOLDING

In eukaryotic systems, many of the proteins that are synthesized in a cell are destined for secretion to the extracellular environment. These proteins are translocated into the ER,[5] where folding takes place before secretion through the Golgi apparatus. The ER contains a wide range of molecular chaperones and folding catalysts, and in addition the proteins that fold here must satisfy a ‘quality-control’ check before being exported. Such a process is particularly important because there seem to be few molecular chaperones outside the cell, although one (clusterin), at least, has recently been discovered. This quality-control mechanism involves a remarkable series of glycosylation and deglycosylation reactions that enables correctly folded proteins to be distinguished from misfolded ones. The importance of these regulatory systems is underlined by recent experiments that suggest that a large fraction of all polypeptide chains synthesized in a cell fail to pass this test and are targeted for degradation. Like the ‘heat shock response’ in the cytoplasm, the ‘unfolded protein response’ in the ER is also stimulated (upregulated) during stress and, as we shall see below, is strongly linked to the avoidance of misfolding diseases [11].

On one (reductive) level, life may be thought of as the co-ordinated activity of proteins, and disease as an imbalance of proteins that adversely affects the quality of life – either through too little of a particular protein being present, or too much of a protein, or a protein being produced or rendered dysfunctional, or produced at the wrong place or the wrong time. Inappropriate folding is one way in which a protein imbalance may arise – the misfolded protein may be nonfunctional or suboptimally functional, or it may be degraded by cellular machinery, or the misfolding may expose epitopes which lead to dysfunctional interactions with other proteins. There are a number of serious diseases which have a common aspect in that they all appear to involve inappropriate folding of a particular protein. These diseases are sometimes lumped together under the heading of the protein misfolding [12,14].

## 12. PROTEIN MISFOLDING DISEASES

Folding and unfolding are the ultimate ways of generating and abolishing specific types of cellular activity. In addition, processes as apparently diverse as translocation across membranes, trafficking, secretion, the immune response and regulation of the cell cycle are directly dependent on folding and unfolding events. Failure to fold correctly, or to remain correctly folded, will therefore give rise to the malfunctioning of living systems and hence to disease. Some of these diseases (such as cystic fibrosis and some types of cancer) result from proteins folding incorrectly and not being able to exercise [7].

In many cases, misfolded proteins are recognised to be undesirable by a group of proteins called heat shock proteins, and consequently directed to protein ubiquitin, which acts as a tag that directs the proteins to proteasomes, where they are degraded into their constituent amino acids. Hence many protein misfolding diseases are characterised by absence of a key protein, as it has been recognised as dysfunctional and eliminated by the cell’s own machinery. Diseases caused by lack of a particular functioning protein, due to its degradation as a consequence of misfolding [8].



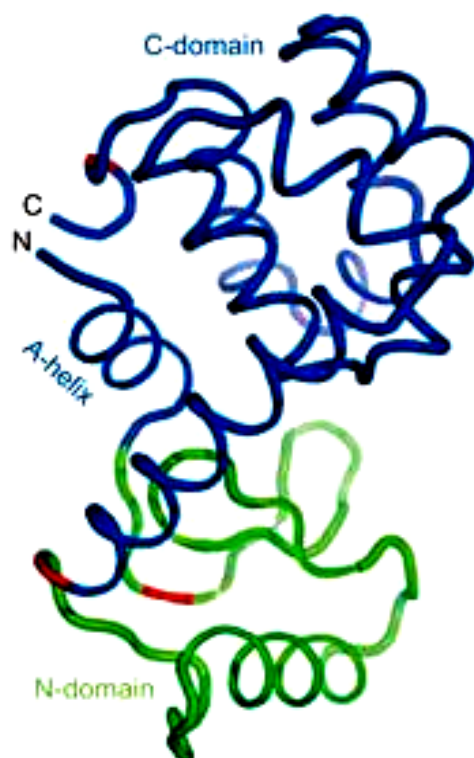


Fig. 4. Misfolded Protein

Diseases caused by Protein aggregation include Alzheimer's disease, Type II diabetes, Parkinson's disease. Protein misfolding appears at least in some cases to be due to mutations (missing or incorrect amino acids) in the protein which destabilise it such that it is more likely to fold incorrectly [4].

### 13. TREATING PROTEIN MISFOLDING

The purpose of studying any human disease is to find ways to treat it. The story of protein folding has not yet led to treatments for the diseases involved, but this could happen within the next decade. The key is to find a small molecule, a drug that can either stabilize the normally folded structure or disrupt the pathway that leads to a misfolded protein. Although many molecular biologists and protein chemists believe this will be quite difficult, others are more optimistic. It is difficult to pinpoint where the search for treatment currently stands, however, one research group has shown that both thyroid hormone and the related compound TIP (2,4,6-triiodophenol) can stabilize transthyretin. Since TIP neither blocks the action of thyroid hormone nor exerts any hormone-like effects of its own, it appears to be a promising treatment for the hereditary disease familial amyloidotic polyneuropathy (FAP), peripheral nerves and other organs are damaged by deposits of amyloid-type protein [1,2,4].

Developing small-molecule therapies is quite straightforward for proteins like transthyretin that naturally bind small molecules, but these therapies are more difficult to apply to proteins that do not have a small-molecule binding site [13].

## 14. CONCLUSION

Folding and unfolding are the ultimate ways of generating and abolishing specific types of cellular activity. Such a process would explain why most of the amyloid (the types of aggregate that can be formed by proteins) diseases are associated with old age, when there is likely to be an increased tendency for proteins to become misfolded or damaged, coupled with a decreased efficiency of the molecular chaperone and unfolded proteins responses. It is ironic that through our success in increasing the life expectancy of the populations of the developed world, we are now seeing the limitations of our proteins and of the regulatory mechanisms that control their behaviour. It is therefore essential that we use our developing understanding of 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.

Molecular chaperones do not themselves increase the rate of individual steps in protein folding; rather, they increase the efficiency of the overall process by reducing the probability of competing reactions, particularly aggregation. However, there are several classes of folding catalyst that accelerate potentially slow steps in the folding process. The most important are peptidylprolyl isomerases, which increase the rate of *cis-trans* isomerization of peptide bonds involving proline residues, and protein disulphide isomerases, which enhance the rate of formation and reorganization of disulphide bonds. Despite these factors, given the enormous complexity and the stochastic nature of the folding process, it would be remarkable if misfolding never occurred. Clear evidence that molecular chaperones are needed to prevent misfolding and its consequences come from the fact that the concentrations of many of these species are substantially increased during cellular stress.

## REFERENCES

- [1] Christopher M. Dobson, R. John Ellis, *The EMBO Journal* 17 (1998) 5251-5254.
- [2] Radouil Tzekov, Linda Stein, Shalesh Kaushal. *Cold Spring Harb Perspect Biol* 3 a007492 (2011) 1-10.
- [3] Feng Zhao, Jian Peng, Jinbo Xu, *Bioinformatics* 26(12) (2010) i310-i317.
- [4] Hironobu Naiki, Yoshitaka Nagai, *J Biochem* 146(6) (2009) 751-756.
- [5] Ashish Arora, Dennis Rinehart, Gabor Szabo, Lukas K., Tamm, *The Journal of Biological Chemistry* 275(3) (2000) 1594-1600.
- [6] Lisa D. Cabrita, Christopher M. Dobson, John Christodoulou, *Current Opinion in Structural Biology* 20 (1) (2010) 33-45.
- [7] Yun-Tzai Lee, Tz-Hsiang Su, Wei-Cheng Lo, Ping-Chiang Lyu, Shih-Che Sue, Circular Permutation Prediction Reveals a Viable Backbone Disconnection for Split Proteins: An Approach in Identifying a New Functional Split Intein *PLoS One* 8, e43820 (2012) 71-12
- [8] Lila M. Gierasch, *Protein Science* 20 (2011) 783-790

- [9] William R Skach, *Nature Structural & Molecular Biology* 16(6) (2009) 606-612.
- [10] Peter L. Freddolino, Christopher B. Harrison, Yanxin Liu, and Klaus Schulten, *Nat Phys.* 6(10) (2011) 751-758.
- [11] Ferenc Orosz, Judit Ovádi, *Bioinformatics* 27(11) (2011) 1449-1454.
- [12] Yien Che Tsai, Allan M. Weissman, *Genes & Cancer* 17 (2010) 764-778.
- [13] Jeung-Hoi Ha, Stewart N. Loh, *Chemistry* 25 18(26) (2012) 7984-7999.
- [14] John H. Van Drie, *Chin J Cancer* 30(2) (2011) 124-137.