Chapter 1
Introduction

Raj Kumar and Bal Ram Singh

The particular field which excites my interest is the division between the living and the non-living, as typified by, say, proteins, viruses, bacteria and the structure of chromosomes. The eventual goal, which is somewhat remote, is the description of these activities in terms of their structure, i.e. the spatial distribution of their constituent atoms, in so far as this may prove possible.

Francis Crick

Keywords Protein Folding • Inverse Protien Folding • Protein Conformation • Protein Flexibility • Protein Dynamics • Protein Toxins • Botulinum Toxins • Protein Science

Every new scientific discovery unleashes a new excitement and fresh concerns: generates several valid and invalid questions. Some people will be in favor of that some oppose. Linus Pauling and Robert Corey published several papers on the inner workings of proteins—molecules regarded as the building blocks of life—which are the key to understanding biology at the molecular level. They might or might not have realized that they are throwing a very important and essential problem to solve for getting the answer of the existence of biological life in this universe. Protein folding problems (PFP) started with the discovery of the structure of the first protein and attracted attention from all kinds of people irrespective of their fields. But this also started a long ongoing quest to answer how these structures form. Over the last 50 years, enormous advancement has been made to understand protein folding mechanisms. But consensus between experimental and theoretical explanation has not reached. Both these approaches are in agreement to some extent for small and simple peptides, but for large and complex proteins any agreement is quite far. PFP has

R. Kumar, B.R. Singh, *Protein Toxins in Modeling Biochemistry*, SpringerBriefs in Biochemistry and Molecular Biology, DOI 10.1007/978-3-319-43540-4_1

three important questions: (a) thermodynamic question: how inter-atomic forces act on an amino acid sequence, (b) Kinetic question: how protein can fold so fast, and (c) computational/technological question: how to predict protein structure. In fact, every protein is unique in its structure and function. Even the homologous proteins are different in their organizations. Part of the problem lies in our approach in designing experiments for studying this phenomenon. We mostly use recombinant protein and study folding-unfolding in the non-native environment using buffer, temperature, chaotropic agents, or protecting osmolytes. We ignore a very important aspect that these are not the natural environment of the protein, and that is the reason why we have a time-lag of denaturation/renaturation process in biochemical labs in comparison to biological lab, the cell. The natural environment of a protein is the cellular environment. When we want to study and get some definite answer then we need to look into this problem in a holistic manner. We need to design our experiment such that we can learn the choreography of this process in its natural milieu, which can be extrapolated to outside environment. Another option for solving PFP is using inverse protein folding (IPF) approach. Here the question is exactly opposite to what we were considering earlier. How the requisite sequence emerges from the functional need for the conformation via evolution or any other method of sequence selection. So, first choose required fold and get the stable fast folding amino acid sequences. In other word, natural end of IFP is the prediction of stable fast folding sequence which folds in situ to biologically useful target conformation/ensemble. A solution to protein folding problem is of enormous intellectual importance, in that it will provide us a "missing link" of information flow from DNA to complete 3D-structure. Not surprisingly, solution to this problem has applications outside the basic research in protein chemistry, such as design of drugs and enzymes.

Whilst occurring quickly, the folding of a protein typically does not happen in one step, as believed previously; it proceeds with several intermediately folding states. Each intermediately folded state is lower in energy than the unfolded state and higher in energy than the completely folded state. These intermediate states, which have key contacts, are important steps in folding that are crucial in directing the steps of folding. One of the important intermediates is termed as molten globule (detailed in Chaps. 2 and 3). In many globular proteins, the molten globule conformation is observed in the early stage of folding. Existence of molten globule in early stage of folding might be a way by which protein can avoid huge conformational space and reduce the time of folding. Because of diversity in a molten globule conformation, it is difficult to get the clear structural detail of this state. It creates confusion too, because in one instance this state forms a flexible native-like tertiary fold (example: cytochrome c and apomyoglobin), while in other instances it forms non-native tertiary structure (example: equine lysozyme). Nevertheless, it is widely accepted that a thermodynamically stable conformation state exists between native and unfolded state. Dry molten globule, wet molten globule and pre-molten globule states are other forms of this state. This important conformation state does not only have structural implications but it has functional effects too. This conformational state plays an important role in various biological processes. The non-native conformations are required in various translocation processes

across a biological membrane. Various molecular chaperones recognize non-native states to fold them properly after the functional need for unfolded structure is over. Various genetic diseases can be caused by the misfolding or non-native structure of translated polypeptides. Because the molten globule is regarded as a non-native structure under physiological conditions, it definitely assumes some role in the above mentioned phenomena in vivo (details in Chap. 3). However, there is a much greater diversity in the non-native conformations of proteins along with the conformations characterized as the molten globule state. That is one of the reasons that PFP problem is getting bigger and bigger. More technological advancements are not helping us. Although technological advancements are helping us in understanding PFP, but they are also unleasing new dimension to this problem.

Growing structural biology evidence is suggesting that protein dynamics and flexibility are more relevant in function than the structure itself. This whole paradigm in which functional state of a protein needs a fixed three-dimensional structure is challenged with the discovery of intrinsically disordered proteins (IDPs). Despite their lack of structure, IDPs are a very large, diverse and functionally important class of proteins. Discovery of IDPs strengthens the following hypothesis by Dr. Matveev: "Native aggregation is regulated aggregation that occurs when interacting protein form new temporary structures through highly specific interactions." According to this hypothesis, cell function can be considered as a transition between two states; resting state and active states. Resting state is a state when protein is natively unfolded states or globular state. In this state cell is transparent. When a cell is activated, secondary structures appear in the natively folded state, which interacts very specifically with the secondary structures of the other proteins and cell becomes turbid. This definition or cellular action fits with the action of IDPs. IDPs can adopt a fixed three-dimensional structure after binding to other macromolecules. To align with the above hypothesis it can also be interpreted that during the transition from resting to the active state of cells protein molecules can transform from IDPs to three-dimensional folded state. Protein conformations in IDPs are flexible enough to facilitate this transition. Because of the inherent structural flexibility IDPs can play an important role in several cellular functions such as protein stability, inhibition, signaling, binding, and regulations. IDPs have been shown to be largely functionally associated with signaling and recognition pathways, cell-cycle regulation, and gene expression. These pathways are also heavily linked to many types of cancers and disease which make IDPs a valuable and challenging therapeutic target.

To study these important aspects of proteins, a model is needed. The selection criteria for a well-defined model would be; (a) it should provide sequence, structural and functional diversity, (b) it should have structural and dynamical flexibility, and (c) a protein class, which covers various aspects of protein chemistry. One class of protein which fits in all the above criteria—Protein Toxins. Toxins are ubiquitous and spread in all three forms (microbial, plant, and animals) of life. They are found in the form of small molecule, simple protein, mixture of proteins and large protein complexes as well. Various animal and bacterial protein toxins provide a structural and sequence diversity, such as the three-finger toxin, A-B toxin. In many cases,

toxins have unique folds and novel combination of domains of known folds. Recombination between toxin genes and sequence divergence has resulted in a wide range of host specificities. This diversity of protein toxins is associated with their structural flexibility. Mostly protein toxins are subjected to host response related stress. Structural flexibility offers a unique ability in adapting to an unfamiliar environment in performing functional actions of these toxins. These toxins affect various cellular functions, such as ion-channel functions, neurotransmitter release, protease activity, aggregation, activation/deactivation, etc. Because of their structural flexibility toxins can exist in several structural conformations such as IDPs and molten globules. These toxins are ubiquitous and because of that, they have unlimited targets. Structural flexibility and diversity in structure and function make toxin family a good candidate to study the evolutionary phenomenon. Since most protein toxins are produced by animal or bacteria and affect different species, they are also a good candidate for examining symbiotic relationships which is the goal of every living species for survival. There are several unanswered questions related to the existence of these molecules—(a) Why these molecules are synthesized, (b) How these toxins find their cellular targets, and (c) Do these molecules play any physiological role for the organism producing such molecules, and (d) how a complex unit of toxins comes together. Careful study and adequate experimental data are needed to answer these questions. On an entirely separate note, these molecules are very useful therapeutic candidates. The difference between a harmful molecule and a beneficial drug is largely the concentration. At therapeutic dose, toxins can be a wonder drug, such as snake venom and botulinum toxin.

This book summarizes unique aspects of protein toxins, mainly concentrating on various aspects of toxins which will be defining its utility in protein chemistry. The book is a sincere effort to establish toxin molecules as a better candidate to be considered as a model of protein chemistry. This is a first attempt of its kind, and we hope to have succeeded in at least some interesting questions, and in shedding some light on interesting possibilities of basic and applied research in protein science.