SPRINGER BRIEFS IN BIOCHEMISTRY AND MOLECULAR BIOLOGY

Raj Kumar
Bal Ram Singh

Protein Toxins in Modeling Biochemistry

Springer

SpringerBriefs in Biochemistry and Molecular Biology

More information about this series at http://www.springer.com/series/10196

Raj Kumar • Bal Ram Singh

Protein Toxins in Modeling Biochemistry

Springer

Raj Kumar
Institute of Advanced Sciences
North Dartmouth, MA, USA

Bal Ram Singh
Institute of Advanced Sciences
North Dartmouth, MA, USA

ISSN 2211-9353 ISSN 2211-9361 (electronic)
SpringerBriefs in Biochemistry and Molecular Biology
ISBN 978-3-319-43538-1 ISBN 978-3-319-43540-4 (eBook)
DOI 10.1007/978-3-319-43540-4

Library of Congress Control Number: 2016951927

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

I have been fortunate to start my graduate research in biochemistry/biophysics with an excellent biochemist/biophysicist. My introduction to research started with basic knowledge of chemistry which helped in my graduate days to connect with biochemistry. Knowledge of basic chemistry helped me to understand biochemistry in better ways. When I started working with Dr. Bal Ram Singh, I did not have any clue what biochemistry is. But, I learned many aspects of biochemistry during my Ph.D. days. I am, of course, most indebted to my research supervisor, Dr. Bal Ram Singh, who has proved to be a good mentor and human being. Supervisor and student are always said to be the analogue of master and apprentice, and I believe Dr. Singh and I complemented this aspect very well.

Also, in building theoretical knowledge, discussion with Dr. Valeri Barsegov (an expert in theoretical biochemistry) helped me very much to clarify several ideas.

Finally, I gratefully recognize the Institute of Advanced Sciences for supporting me with financial and research resources. This book is a good opportunity for me to enhance my knowledge in this area and put forward my thoughts/ideas about protein chemistry of toxins.

Professor Singh and I conceptualized this book a few years ago. We wanted to turn the table around and promote toxin as a congenial molecule rather than a molecule we needed to be scared off. We wanted to promote the unique characteristics of this molecule which is rare in other classes of protein family. A very important perspective about toxin, which we wanted to highlight, through this book, is flexibility and its utility in evolution, in general.

The effort has been made to keep each chapter of this book largely simple, self-contained, and with acceptable arguments, wherever necessary. This is meant to be for the ease of following the subject by the reader. References are carefully selected for each chapter. We have described protein folding (in general along with various theories), one of the longest ongoing problems in science (in Chap. 2). Other chapters are dedicated to the existence of toxin in different unique conformational states, intrinsically disordered proteins and molten globule, and biological meanings of

these conformations. The existence of toxin in these forms explains their adaptability in the host environment. Finally, we discuss the evolutionary characteristics of toxin molecules.

Although most of the information in this book is available in the database, there is no concise information available specially relating to the toxins. As a result, scientists have mostly negated or paid less attention to several unique molecular characteristics of the toxins. Toxin as a molecule, with its uniqueness, has a lot to offer for protein chemists, in particular, and scientists from across the fields, in general.

I hope readers will find the book useful, and it will trigger other provocative thoughts of novel avenues of looking at molecular evolution.

Dartmouth, MA Raj Kumar
June 2016

Contents

List of Figures

List of Tables

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.

Chapter 2
Introduction to Protein Folding

Raj Kumar and Bal Ram Singh

Keywords Protein folding • Folding nucleus • Energy landscape • Ramachandran plot • Molten globule

2.1 Introduction

It is universally accepted fact that organizations of proteins are not simple, repetitive units, like DNA, but consist of compact folds which are difficult to predict. The protein folding problem (PFP) is one of the longest elusive problems, drawing attention of many scientists from several disciplines. In regards to the PFP there are several questions that need to be considered: (a) Is there any code that exists to guide translation from the sequence of amino acids to a completely folded structure; (b) What are the molecular factors that direct and control rate of the folding process; (c) What are the factors that stabilize a native structure; and (d) What is the evolution of folding nucleus, if it exists, and folding kinetics? Coded language of protein folding and functions lies in its sequence and conformations. Folding of a protein occurs through various pathways. Understanding these pathways is an interesting challenge because of the fact that these are determined through the sequence of hundreds of amino acids and various intermolecular forces. There is a definite path (or at least a set up pathways) proteins follow, as a random trials through all conformations would cost an enormous amount of time, approximately 10^{100} years to fold a simple 100 residue protein (Anfinsen 1973; Bryngelson et al. 1995).

Earlier, protein folding was believed as a cooperative process: simple two state process involving transition between folded and unfolded state. Privalov (1989) evaluated the criteria of cooperative transition in several proteins. This model of protein leads to theories like polymer collapse (Ptitsyn et al. 1968), side chain packing (Shakhnovich and Finkelstein 1989), directional or three-body interactions (Hao and Scheraga 1998), special folding pathways (Dill 1985), or statistical-mechanical theories (Ramanathan and Shakhnovich 1994; Sfatos et al. 1993). But discoveries of stable intermediates vindicate the idea of cooperative two step transition (Baldwin and Rose 1999). Shallow transition characteristics were observed for many proteins, which leads us to hypothesize that neither the native state nor the unfolded state can be defined as a single conformational state, rather both these states are a collection of several conformational states. For simplification, Protein folding has four general

R. Kumar, B.R. Singh, *Protein Toxins in Modeling Biochemistry*, SpringerBriefs in Biochemistry and Molecular Biology, DOI 10.1007/978-3-319-43540-4_2

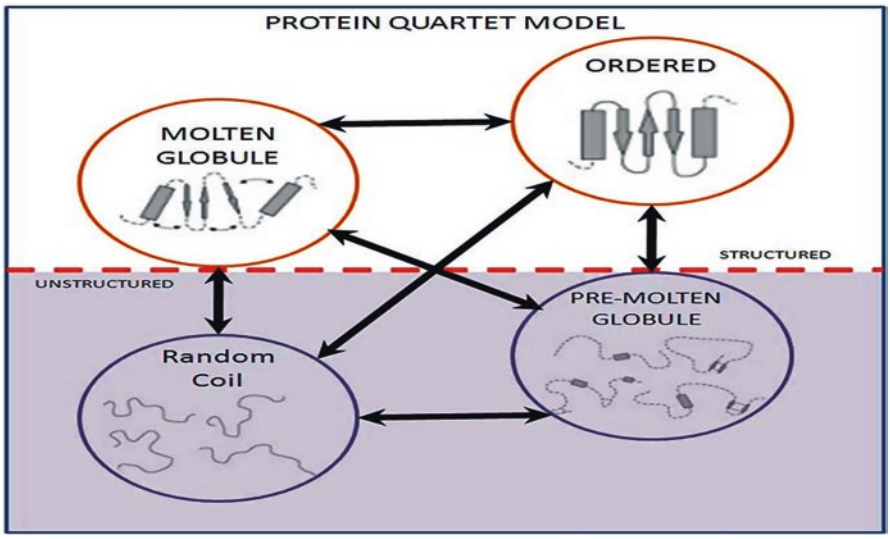

Fig. 2.1 Proteins and protein regions may exist in at least four different states of structural organization: ordered, molten globule, pre-molten globule, and coil-like. Protein function is associated with any of these distinct states or with transitions between them. Entire proteins or protein regions in the pre-molten globule state or coil-like state display no unique tertiary structure. These Intrinsically Disordered Proteins (IDPs) and Intrinsically Disordered Regions (IDRs) are involved in key biological processes including transcription regulation, cell cycle control, recognition, and signaling. Reproduced with Permission from (http://en.bioinformatyk.eu/contest-articles/prediction-and-application-of-intrinsically-disordered-regions-in-practice.html)

structures, according to the Protein Quartet Model—unfolded (random coil), pre-molten globule, molten globule, and ordered (Fig. 2.1; Gorka 2011).

Research has been directed not only to understand the native and unfolded states; acceptance is growing to understand the conformation of intermediates, such as molten globule. In addition, Levinthal's paradox exposes us to the basic conundrum wherein an ordered state emerges spontaneously from an unordered state. The unfolded/unordered state is a generic term, and is very diverse to have specific definitions. More precisely, the denatured chain or unfolded state behaves like a statistical coil and can be described as a rotational isomeric state model (Flory and Volkenstein 1969). However, it is possible to measure the degree of expansion and contraction, by their radius of gyration, R_G (Eq. 2.1).

$$R_G = \sqrt{\frac{\sum_{i=1}^{N} R_{Gi}^2}{N}} \tag{2.1}$$

Where R_{Gi} is the distance of atom i from the center of gravity and N is the number of atoms in the molecule (Fleming and Rose 2005). For a statistical coil polymer with excluded volume, the radius of gyration is (Flory and Volkenstein 1969).

$$R_G = R_0 . n^{\nu} \tag{2.2}$$

Where R_0 is a constant that depends on the intrinsic chain stiffness, n is the number of residues, and ν is the exponent of interest that depends on solvent quality, range from 0.33 (for compact structure) to 0.60 (for denatured protein). Experimentally it is demonstrated that mean radii is consistent with the theory (Tanford 1968). In summary, individual molecules can exist as unfolded in denaturing conditions, but can adopt a unique native condition under favorable folding conditions.

However, the protein folding story is not just about native state folding or ensemble of conformations at native state, it has another dimension too. Contrary to earlier belief, a protein does not need a complete compact structure for its function. Folding of proteins is not just a theoretical idea, it has a biological implications. It also raises a fundamental question, is folding necessary for biological functions? Several proteins exist without a closed packed structure, such as intrinsically disorder proteins (IDPs, discussed in Chap. 3). IDPs are proteins with significant amount of disorder present in their native state. Existence of IDPs and complexity of an organism has a relationship, more complex the organism more IDPs involved in their cellular function. Eukaryotes have more than 50 % of proteins which are classified as IDPs. Instead of closed packing, flexibility is the driving force of complexity and ultimately the evolution. Concept of native aggregation, in which protein secondary structure interacts specifically with the secondary structure of another protein (Matveev 2010), forced us to rethink if folding is an ultimate requirement for biological function.

Concept of IDPs and native aggregation gave PFP problem an entirely different dimension and challenged conventional view of protein folding in which native state of a protein is a well packed functional structure. This confusion arises because of existing disconnect between the understanding of *in-vitro* and *in-vivo* processes. Nevertheless, folded state is a well defined state, and search of protein molecules for this state out of several possible states itself generates a scientific excitement. In subsequent sections, we will focus on understanding protein molecules in terms of their structure and folding behavior.

2.2 Classification of Protein Family

Protein classification is as complex as its structure and folding. Although it can be simply classified as simple proteins: if they yield only amino acids after digestion with hydrolases (Albumins, Globulins, Glutelins, Prolamines etc.), and conjugate proteins: if they yield amino acids and additional products after digestion with hydrolases (Lipoproteins, chromatoproteins, metalloproteins, nucleoproteins etc.).

Another classification could be on the nutritional basis: complete protein, if they supply all the essential amino acids and incomplete protein, if they are deficient in one or more essential amino acids. But more sophisticated and scientific classification will be based on family of proteins. There are various kinds of proteins which can be divided in structural, transport, storage, enzyme, etc. (Table 2.1).

Table 2.1 Classification of proteins

Class of protein	Function in the body	Examples
Structural	Provide structural components	Collagens, Keratin, Elastin
Contractile	Move muscles	Myosin, Actin
Transport	Carry substances from one part of the body to other part; one part of cells to other part of cells	Hemoglobin, Myoglobin and Lipoproteins
Storage	Store nutrients	Casein, Ovalbumin, Ferritin, Globulines, Prolamines, Glutelins
Hormone/regulatory	Regulate body metabolism and nervous system	Insuline, Adrenalin, Thyroxin
Enzyme	Catalyze biochemical reactions in the cells	Trypsin, Pepsin, ATPases, Cytochrome C, Lactase, Ribonuclease, Fumarase
Protection	Recognize and destroy foreign substances	Immunoglobins (different classes), Fibrins
Genetic	Carry, store and protect genes	Nucleoproteins, Histones
Toxic	Mostly their functions are to kill cells	P53, Lectin (endogeneous). Ricin, Abrin, SEB, Botulinum Neurotoxin (Exogeneous bacterial proteins)

Other classes of proteins could be fibrous proteins (long thread like molecule), and globular proteins (spheroid shape molecules).

Another common way of classification is to divide all proteins or protein domains (defined in the next section) into five major groups; all α, all β, α/β, α + β and a fifth group of proteins which do not fit in any other categories. The α/β and α + β proteins differ in that the α/β proteins are mainly built up of parallel strands connected by helices (βαβ motifs), while the α + β proteins have strands and helices connected in a more irregular fashion. Within each group there are many folds.

2.3 Types of Fold

Protein fold is defined as a structural arrangement of secondary structure in three dimensional space. For example, Rossman Fold (named after the discoverer M. G. Rossman; Fig. 2.2) of co-enzyme dehydrogenase. Similarly, there are other folds like Jelly roll fold, TIM (Triosephosphate isomerase) fold known to us. According to SCOP (Structural Classification of Proteins) there are 1393 folds and CATH (Class, Architecture, Topology and Homologous Family) has 1282 folds.

SCOP database is based on a manual classification of the folds. The levels of the hierarchy are classes, fold, superfamilies and families. The domains within a family are supposed to be homologous, having a common ancestor from which they diverged. Proteins within one superfamily have the same fold and related function,

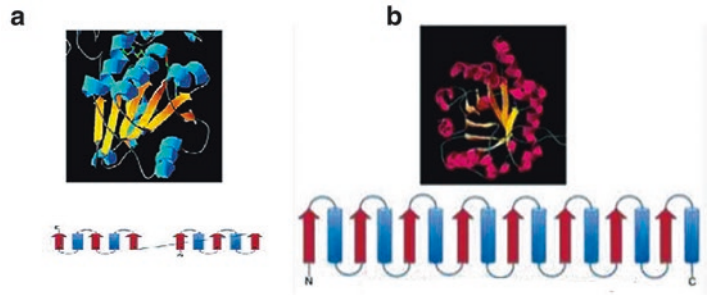

Fig. 2.2 Rossman fold is composed of seven parallel β-sheets, first two strands are connected by α-helix (**a**). Rossman fold is frequently occurring domain in nucleotide binding proteins. Triosephosphate isomerase (TIM) fold consists of eight α-helix and eight parallel β-strands that alternate along peptide backbone (**b**)

possibly having a common ancestor, but they can differ in sequence and function. At the next level, the fold, the proteins have the same topology, but there is no evidence of evolutionary relationships. Finally, the protein in the fold can have similar secondary structural organizations such as all alpha, all beta, etc.

CATH is a similar structure database, which uses automated procedures for classification. It is similar to SCOP but the architectural level is not defined as in the case of SCOP. Proteins with the same overall secondary structure, but with different connections are termed as the same architecture. Because of this difference there are some differences in grouping of the proteins.

There are two more databases which are not hierarchic but are based on structural comparison of the known structure: FSSP (Fold classification based on Structure-Structure alignment of Proteins) and Enterz database.

2.4 Domains of Proteins

A domain of proteins can be defined as the basic building block of a protein structure, although many proteins only contain a single domain. Some proteins exist in multiple domains. Even though it is not necessary to have a specific function for each domain, certain protein domains have some clearly defined functions associated with them, like the Rossmann-fold domain, also called coenzyme-binding domain or botulinum toxin has endopeptidase domain, translocation domain and binding domain. Such domains often "carry" their function with them when they get inserted into different proteins during evolution. A domain may be characterized by the following:

1. A spatially separated unit of the protein structure.
2. Sequence and/or structural resemblance to another protein structure or region.
3. Specific function associated with that segment of the protein.

2.5 Protein Folding

Proteins are the essential part of an organism and virtually take part in all the functions within a cell. Each protein has its unique sequence and that sequence is encoded in its gene. Encoded genes are first transcribed into pre-messenger RNA, which is processed into mature messenger RNA (mRNA). mRNA has all the information, including post-transcriptional modification and is used as a template for protein synthesis by the ribosomes. Proteins are always biosynthesized from N-terminus to C-terminus. After translation, a protein exists as unfolded polypeptide lacking any three-dimensional structure. Various interactions of amino acids lead to the development of a well-defined three-dimensional structure, known as the native state. The process of formation of a well-defined three-dimensional structure from an unfolded structure is known as folding. A correct folding is essential for the proper function of a protein. Failure to fold produces an inactive protein, which may be toxic. Several neurodegenerative diseases are believed to result from the accumulation of misfolded protein.

An essential question in protein folding studies is how an unstructured peptide folds into its unique active state and what is the governing mechanism of this process. To get to this state a protein has to go through an astronomically large number of possible conformational states (Levinthal's Paradox). The presence of folding pathways in which some intermediates are well populated could be the answer to this paradox and could help us to select these conformational intermediates. A typical small protein can fold reversibly, in vitro, undergoing a first order transition between the folded and unfolded state. Not only finding a native state is desired, exquisite side chain packing in the protein core, binding sites, and catalytic sites are important for molecular recognition and protein function. Contrary to this view, an intermediate state in protein folding can adopt native state and have very specific interactions even in the absence of side-chain close-packing. Proteins are large biomolecules consisting of one or more chains of amino acids. In spite of their enormity of size in terms of number of atoms the overall size of proteins is a few nanometers in diameter. Because of folding the proteins are often termed as nano-machines. Proteins have mainly four levels of structure: primary, secondary, tertiary and quaternary (Fig. 2.3). In addition to these levels of structure, proteins change its structure in response to the environment as well, to perform their functions. These transitions in structure, which is mainly associated with tertiary and quaternary structure (sometimes there is a significant change in secondary structure as well), coupled with side chains rearrangements are referred as conformational changes. Such conformational changes are typically associated with binding to substrate molecule, thermal vibrations of solvent molecules or due to interaction with other molecules (molecular crowding).

A few pertinent open questions regarding protein structures and foldings are as follows. (a) What is the correlation between sequence and three-dimensional folding? (b) What is the role of intermediates in protein folding? (c) How does a protein discard undesired conformations? (d) Intermediates are markers of protein folding? (e) Does disorder in protein structures play a role in protein function?, and (f) What is the role of the microenvironment and internal structure of protein in folding a functional structure of a protein?

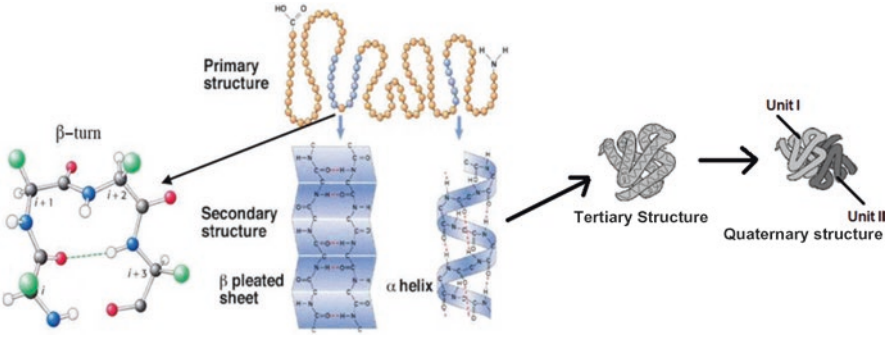

Fig. 2.3 Different structural organization of a protein

2.5.1 Conventional Views on Protein Folding

Some proteins spontaneously fold into the correct 3D structure under a reasonably wide range of conditions. This view is not universally true.

- Many proteins only adopt their functional structure in very specific environments, e.g., membrane proteins.
- Some proteins require assistance to fold the correct way. Molecules known as chaperones assist these polypeptides to fold incorrect conformations.
- Protein folding is a delicate and complex process. But misfolding can occur even with right primary sequence, such as prions in mad cow disease. Many neurodegenerative diseases are associated with an accumulation of such misfolded proteins.

Protein folding has been a widely studied and debated topic in biology for decades. The challenge of understanding the pathway producing a folded protein has been daunting as each protein begins as a single strand of hundreds/thousands of amino acids and relies on a series of intramolecular forces (hydrogen bonds, hydrophobic interactions, interactions between charged residues, covalent links, backbone angle preferences, chain entropy) and potentially molecular chaperone proteins to reach the final native state within a timescale of as low as microseconds (Hartl et al. 2011). An understanding of protein folding is essential to better understand how a cell operates with such an abundance of (about 300–400 g/L concentration) proteins (Hartl et al. 2011) and nature of proteins which control biological activities such as cellular trafficking, cell growth, differentiation, enzymatic activity, and even acting as a chaperone for assisting other proteins fold correctly (Dobson 2001; Hartl et al. 2011).

There is a definite path protein follow as a random search through all conformations would cost an enormous amount of time, approximately 10^{10} years to fold a simple 100 residue protein (Bryngelson et al. 1995). Newly synthesized peptides from the ribosomes are in a high state of energy due to the flexibility and lack of any intramolecular structure and tend to migrate to a lower energy state through a series of intramolecular or intermolecular interactions. This decrease in energy can include the formation of secondary structure (α-helices or β-sheets), formation of oligomers or the formation of amorphous aggregates or amyloid fibrils (Hartl et al. 2011).

This protein energy funnel contains many local energy minima which allow for rapid interconversion with only small changes in the energy levels (Bryngelson et al. 1995).

Typically, proteins would tend to avoid any unfolding or disorder, since this would result in an increase in aggregation or misfolding. Increasing the stability of the protein to prevent this would increase the rigidity of the structure. With these issues in mind and the fact that there are many examples of natively unstructured proteins, proteins with disordered fragments, and proteins which adopt a molten globule state, there has to be some significant benefit to these proteins to remain flexible.

2.5.2 Possible Mechanisms of Protein Folding

In 1950, Anfinsen proposed that the amino acid sequence itself can determine the three-dimensional structure of the protein. It would be very useful to be able to predict the structure of a protein from its primary sequence.

1. Predicting the folding and stability of protein will be easy.
2. A better understanding of folding helps to understand better about the biological functions of proteins.
3. Instead of using expensive techniques, such as NMR and X-ray crystallography, one can easily predict the protein behavior just by knowing the sequence.
4. Synthesize novel proteins for different industrial and research purposes.

However, the idea of predicting 3D structures of proteins from the primary structure holds true to some extent only for small globular proteins at best. Use of chaperones for protein folding makes this issue more complicated. Folding of proteins in general and large proteins, in particular involves several intermediates. Frauenfelder and Wolynes (1994) state that the final stages of folding will depend on the specific sequence of amino acids, whereas earlier folding stages/intermediates should be mostly insensitive to details of the sequence. Although mutation of native structure is more or less robust and has less or no impact on the structure, the enzymatic activity does not have similar mutational freedom as even single mutation can reduce or disrupt the activity.

There are various concepts that attempt to explain how exactly a protein undergoes folding and organization from the unfolded or random coil state to the natively folded state. Logically protein folding should not be random and unbiased searching of all possible conformations; it should follow some definite pathways. First proposed mechanism—nucleation-growth mechanism—states that tertiary structure propagates rapidly from an initial nucleus of local secondary structures. Second proposed mechanism evolved after the discovery of intermediates of the proteins. This leads to a framework model mechanism—diffusion model, in which secondary structures form first, then the collapse of secondary structure yields the native conformation. In the framework model, formation of large amount of secondary structure proceeds from a denatured state. Another mechanism, and the most accepted mechanism, is the hydrophobic collapse model, which says hydrophobic collapse drives compactness of protein to native form (Fig. 2.4). The hydrophobic driving

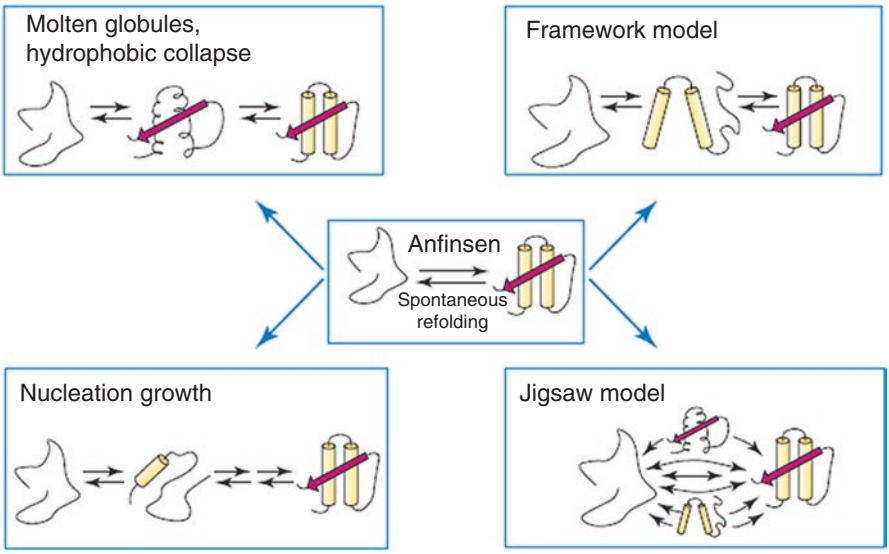

Fig. 2.4 Models of protein folding (Reproduced with permission, Radford 2000)

forces coupled with expulsion of water from the nonpolar residues, allows the formation of secondary structure due to the collapse. Finally in a jigsaw model suggesting that folding of every protein is unique and can follow a different or mixed path.

As discussed above, theory of protein folding is based on four models. Earlier it was believed that the presence of tertiary structure is essential for the formation of stable secondary structure (Epand and Scheraga 1968). Although the idea, provided by Epand and co-workers, does not true in totality as hydrophobic and other interactions stabilize the secondary structure (de Prat et al. 1995), but the study on helical peptides provided a different perspective. When the secondary structure is stable in isolation, it tends to adopt α-helical and β-turn structure and less likely to form β-sheet structures.

The starting point of nucleation model is the concept of a folding nucleus (FN), which is defined by two conditions: (a) formation of a nucleus or set of contacts that constitutes the nucleus, leading to a very fast subsequent folding; (b) the presence of a pattern of contacts related to the folding nucleus in the folding conformations (native contacts are relatively small), leading to a very fast subsequent folding. The defined nucleus is common to all the possible conformations of the protein. Interestingly, the folding nucleus has the cooperative character of the folding transition (Hao and Scheraga 1994; Chan et al. 2004; Faisca et al. 2004; Faisca and Plaxco 2006; Prieto and Rey 2007). Mathematically, folding nucleus can be defined as a set of most probable native contacts which are separated from the native conformation as far away as possible, and yet fold with high folding probability, $P_{fold} \geq 0.9$. The concept of folding probability was coined by Du et al. (1997). In practical term, P_{fold} is defined as a conformational state in which a particular conformation folds before it unfolds. So, $P_{fold} > 0.5$ of a conformational state indicates that this conformational state is likely to fold first. Similarly, $P_{fold} < 0.5$ is likely to unfold first than to fold, $P_{fold} = 0.5$ can go in either direction.

The multiple nuclei model was also proposed by Klimov and Thirumalai (1998). In this model, the formation of the folding nucleus is the rate limiting step, which is identified as the folding transition state (TS). This model argues involvement of multiple nucleus. Depending on the initial conformations many nuclei can be involved. Instead of having one transition state, an ensemble of transition states is possible (TSE; Klimov and Thirumalai 1998). In a single folding nucleus nonlocal contacts are predominant, whereas in the multiple nuclei model both local and non-local contacts are present but local contacts are predominant. Using psi-value analysis, Sosnick et al. (1994) showed that there are two classes of TS heterogeneity, a single TS nucleus (as in the case of Ubiquitin) and multiple TSE corresponding to multiple FN (dimeric GCN4 and titian I27). Thus, it is possible that a group of contacts is required for a single nucleus model or having a single nucleus with microscopic heterogeneity. Since FN involves non-native interactions, these interactions affect the whole process in the TS. The relation between non-native interactions and the nucleus mechanism was explored by Shakhnovich (1998) and Li et al. (2000). They examined the FN with the fraction of native contacts $Q=0.41$ in a single nucleus with microscopic heterogeneity. Interestingly, 45 % of the FN contacts are non-native and predominantly non-local. The transition state resembles a distorted form of the native structure, with the least distorted part being loosely defined as the nucleus, and the distortion tends to increase with increasing distance from the nucleus (Daggett 2002). Studies also indicated that the native geometry determines the distribution of the FN along the protein chain, but the specific location of FN is modulated by protein sequence. In brief, nucleation condensation mechanism invokes both the hydrophobic collapse and framework mechanisms. This mechanism invokes the long range and other native hydrophobic interactions in the transition states to stabilize the secondary structures. Although the nucleation-condensation mechanism was introduced to help explain mainly two state folding, it is not limited to a single domain proteins. Although protein appears to fold in two-state folding, but at the atomic level this may not be the case. For example, unfolding of chymotrypsin inhibitor (CI2; Fig. 2.5a) at 100 C using MD simulation. Denatured state of this protein had residual structure and hydrophobic clusters near to the central part of the protein. Residual structure and hydrophobic clusters could be nucleation sites, and collapse and condensation can occur about these sites (Fig. 2.5b) (Daggett and Frenst 2003). Another example, Barnase is two semi-autonomous domains (Fig. 2.5c). Its folding-unfolding pathway has been studied by MD (molecular dynamics), NMR and other experimental techniques. In its denatured state Barnase has a residual structure, which helps to set up loose, native topology within the denatured state. The transition state identified from the simulation is in quantitative agreement with the experimental phi-values (Fig. 2.5d).

MD simulation analysis of CI2 demonstated that secondary and tertiary structure collapse can proceed together, nucleation-condensation mechanism suggesting hydrophobic collapse and nucleation can happen together. Whereas MD simulation of barnase showed that every secondary structure is actively aiding the formation of other secondary structures via side-chain interactions. This phenomenon is called ‘contact-assisted’ secondary structure formation.

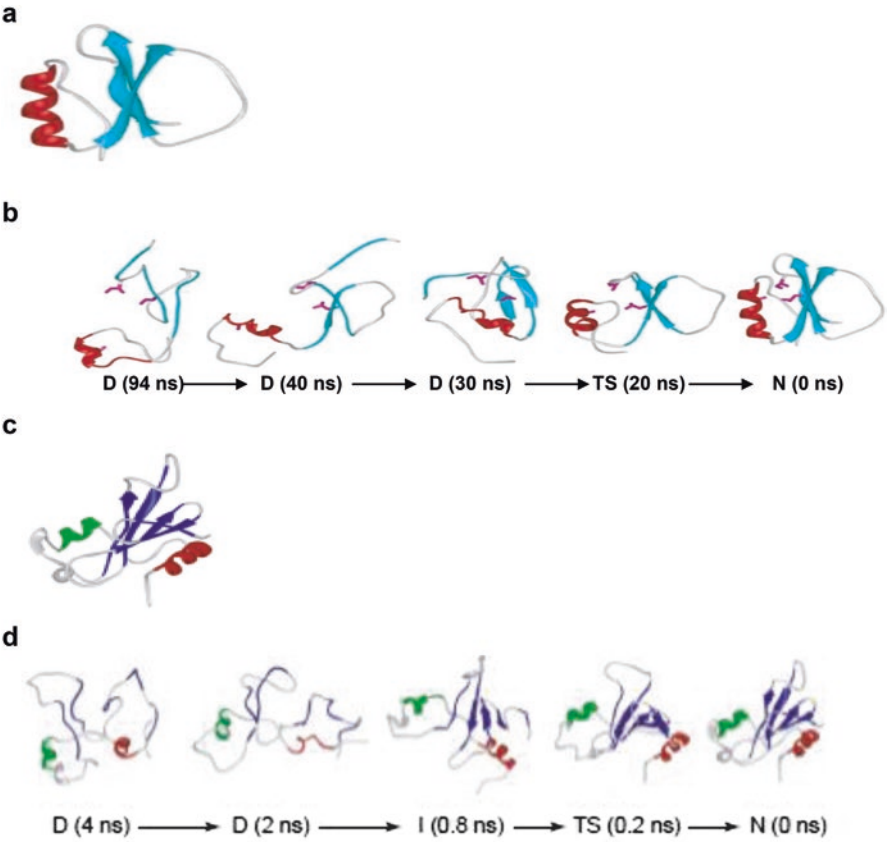

Fig. 2.5 Folding of chymotrypsin inhibitor 2 (CI2). Crystal structure of CI2 (**a**). Structures of CI2 at different time point of thermal unfolding simulation (100 °C) of the folded state to unfolded state. *D* is denatured state, *TS* is transition state, and *N* is native state (**b**). NMR structure of barnase (**c**). Snapshots of unfolding of barnase taken from a 225 C simulation in water (**d**). (Reproduced with permission, Daggett and Frenst 2003)

Although Barnase is small protein, it is reasonable to assume that other larger proteins will follow a similar pathway. Scientists are not sure of success in search of a unifying mechanism. But if the secondary structure is sufficiently stable in the absence of significant tertiary interactions, folding can proceed in a more stepwise, hierarchical manner. It is possible that the diffusion–collision mechanism (hydrophobic collapse and framework) combined with nucleation condensation model (nucleation model) give rise to a stable secondary structure. So the coupling between secondary and tertiary structures is important for folding. If this is weak, then it favors framework model, if strong, then it leads to a stable transition state. In other words, two state folding depends on the particular primary structure and balance of various forces.

2.5.3 *Prediction of Folding from Sequence*

The inverse protein folding problem is most often approached by seeking the sequence that is similar to the sequence of a protein whose structure is known. If a sequence-folding relation is known then it will be easier to predict the fold of protein by knowing the sequence. But the problem is not as simple as it is stated. If the sequence-folding relation is known then it can be inferred that the protein of unknown structure adopts a fold similar to the protein of known structure. Accurate *de novo* prediction of 3D structure from sequence would rapidly enhance overall knowledge of protein chemistry. This fundamental question can be approached by starting a much simpler question: What is required for determination of particular secondary structure; α-helix, β-sheets, turns or coil? Residues such as alanine, glutamate, leucine tend to be present in alpha-helix, whereas valine and isoleucine tends to be part of beta-sheets or strands. Glycine, asparagine, and proline have a propensity for being in turns. Proline is also termed as helix breaker.

Table 2.2 is based on Chou-Fasman rule, which is the most commonly used algorithm. This rule measured frequencies at which each amino acid appeared in particular types of secondary structure. This rule assigns the propensity of amino acids P(a), P(b), and P(turn) of α-helix, β-sheets, and turn, respectively. Also, assign four turn parameters based on the frequency at which they were observed in the first, second, third or fourth position of a beta turn; f(i), f(i + 1), f(i + 2), and f(i + 3), respectively. First this rule identifies helix and sheet nuclei, and then applies a set of rules to determine if the particular amino acids are sufficient to nucleate a region of α-helix or β-sheet.

For helix: 4 out of 6 amino acids with $P(a) > 1$

- extend the nucleus in each direction until reach four amino acids in a row with $P(a) < 1$.
- for each of these regions, add up all the P(a) and all the P(b) values.
- if the total P(a) is larger than the total of P(b) and the run is more than five amino acids then it is predicted to be alpha helix.

For sheet: 4 out of 6 amino acids with $P(b) > 1$ (or use 3 out of 5)

- extend the nucleus in each direction until reach four amino in a row with $P(b) < 1$.
- for each of these regions, add up all the P(a) and all the P(b) values.
- if the total P(b) is larger than the total P(a), the run is more than five amino acids long, and the average $P(b) > 1$ then it is predicted to be beta sheet.

For beta turn:
P(t) for amino acid at position $i = f(i) + f(i + 1) + f(i + 2) + f(i + 3)$
For predicting a beta-turn at position i if the following criteria are met:

- the calculate $P(t) > 0.5$,
- the average P(t) for amino acids i to $i + 3$ is > 0.1,
- the sum of the P(t) values for amino acids i to $i + 3$ is larger that the sum of P(a) and P(b) values.

Table 2.2 Relative frequencies of residues in secondary structures

Name	P(a)	P(b)	P(turn)	f(i)	f(i+1)	f(i+2)	f(i+3)
Alanine	1.42	0.83	0.66	0.060	0.076	0.035	0.058
Arginine	0.98	0.93	0.95	0.070	0.106	0.099	0.085
Aspartic acid	1.01	0.54	0.146	0.147	0.110	0.179	0.081
Asparagine	0.67	0.89	1.56	0.161	0.083	0.191	0.091
Cysteine	0.70	1.19	1.19	0.149	0.050	0.117	0.128
Glutamic acid	1.51	0.37	0.74	0.056	0.060	0.077	0.064
Glutamine	1.11	1.10	0.98	0.074	0.098	0.037	0.098
Glycine	0.57	0.75	1.56	0.102	0.085	0.190	0.152
Histidine	1.00	0.87	0.95	0.140	0.047	0.093	0.054
Isoleucine	1.08	1.60	0.47	0.043	0.034	0.013	0.056
Leucine	1.21	1.30	0.59	0.061	0.025	0.036	0.070
Lysine	1.14	0.74	1.01	0.055	0.115	0.072	0.095
Methionine	1.45	1.05	0.60	0.068	0.082	0.014	0.055
Phenylalanine	1.13	1.38	0.60	0.059	0.041	0.065	0.065
Proline	0.57	0.55	1.52	0.102	0.301	0.034	0.068
Serine	0.77	0.75	1.43	0.120	0.139	0.125	0.106
Threonine	0.83	1.19	0.96	0.086	0.108	0.065	0.079
Tryptophan	1.08	1.37	0.96	0.077	0.013	0.064	0.167
Tyrosine	0.69	1.47	1.14	0.082	0.065	0.114	0.125
Valine	1.06	1.70	0.50	0.062	0.048	0.028	0.053

Reproduced with permission, Chou and Fasman (1978)

Depending on the protein this rule predicts secondary structure with 50–85 % accuracy, with average accuracy ~ 70 %.

Another commonly used algorithm is GOR (Garnier, Osguthorpe and Robson). This algorithm uses a 17 amino acids window to predict secondary structure. The rationale behind this is that the experiment showed each amino acid has a significant effect on the conformation of amino acids up to eight positions in front or behind it. This uses a 17×20 matrix to calculate the propensity of each amino acid within the 17 a.a. windows. Accuracy is about 65 %.

Many other similar programs for modeling secondary structures and motifs are available, such as SIMPA96 (Similarity Peptide Analysis; Levin 1997), DSC (Discrimination of Secondary Class; King and Stenberg 1996), DPM (Double Prediction Method; Deleage and Roux 1987), PHD (neural network) (Rost et al. 1994).

Challenge of sequence-to-structure is enormous and remains unresolved. Computational method that uses fragment libraries or other conformational search followed by energy optimization and molecular dynamic simulation has been very successful at predicting the 3D structures of smaller proteins (Bradley et al. 2005; Simons et al. 1999; DeBartolo et al. 2009; Yang et al. 2007). Although various computer simulation techniques, such as molecular dynamics, Monte Carlo simulation, have come a long way to simulate protein dynamics, but these techniques have yet to achieve folding of extended polypeptides to the native fold.

2.6 Theories of Protein Folding

Among all biological molecules, proteins are the workhorses and important biomolecules for living organisms. The linear chain of amino acids becomes functional only when it folds into a three-dimensional structure. Although it is possible to decode a linear sequence of amino acids, three-dimensional structure predictions is not possible. Like structure, folding of a protein molecule is a very complex process and very specific, which is a central problem in biophysics or molecular biology. As mentioned above, various concepts had been established, such as nucleation-collapse (Guo and Thirumalai 1995), thermodynamics and kinetic intermediate (partition) theory (Guo et al. 1992; Peng et al. 2011; Stigler et al. 2011), minimum energy compete structure (Camacho and Thirumalai 1993; Lin and Zewail 2012), and scaling of folding rates as a function of protein size (Thirumalai 1995). A thought provoking experiment was performed in 1960 by Nobel Laureate Christian Anfinsen and his colleagues. They used a 124 a.a. long protein called ribonuclease. Using this protein they were able to demonstrate that folding-unfolding transition processes are completely reversible. They concluded that the folding information of a protein is present in their amino acid sequence. Foldability of a protein varies from structure to structure, but the specificity of structure arises from the heterogeneity of the protein chain. The time scale of folding, complexity of structure and interactions, and different specific possible path for protein molecules makes folding a difficult problem to study by experimental as well as computational methods.

2.6.1 Ramachandran Plot

In a polypeptide the main chain N-Cα and Cα-C bonds are relatively free to rotate. These rotations are represented by the torsion angles phi and psi, respectively. Ramachandran et al. (1963) used a computer model of small polypeptides to systematically vary ϕ and ψ with the objective of finding stable conformations. For each conformation, the structure was examined for close contacts between atoms. Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii. Therefore, phi and psi angles which cause the spheres to collide correspond to satirically disallowed conformations of the polypeptide backbone.

Ramachandran plot is a two dimensional representation of torsion angle ϕ (angle between N(i − 1), C(i), Cα(i), and N(i)) and ψ (angle between C(i), Cα(i), N(i), and C(i + 1)). All the torsion angles are not possible along the peptide backbone due to steric hinderances/interferences of the side chains. Plotting the torsional angles of each residue in a peptide bond defines the allowed angle in which adjacent residues can be arranged. This exercise discards many angle combinations and sterically stressed conformations between neighboring residues. All the residues (except alanine and proline) can occupy one of the three major "allowed" regions (Fig. 2.6). α and β-region combinations give rise to α-helices and β-strands, respectively.

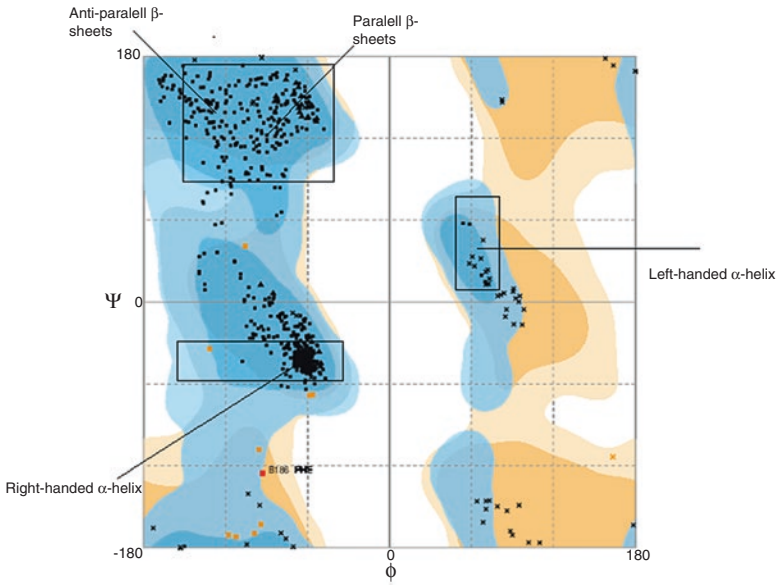

Fig. 2.6 Ramachandran plot of Human UDP-galactose polymerase (PDB ID: 1EK6). Combination of ϕ and ψ angles of different secondary structures is plotted

These regions are "allowed" in the sense that when the peptide atoms are given standard radii they do not collide. An additional region, sometime called bridge region becomes allowed if the atoms are given smaller radii that present the smallest values that could be considered plausible.

- Quadrant I shows a region where some conformations are allowed. This is where rare left-handed alpha helices lie.
- Quadrant II shows the biggest region in the graph. This region has the most favorable conformations of atoms. It shows the sterically allowed conformations for beta strands.
- Quadrant III shows the next biggest region in the graph. This is where right-handed alpha helices lie.
- Quadrant IV has almost no outlined region. This conformation (ψ around −180 to 0°, φ around 0 to 180°) is disfavored due to steric clash.

Exceptions from this principle of clustering around based on torsion angles are glycine and proline. Glycine does not have a complex side chain, due to which it has high flexibility in a peptide chain. That is why Glycine is often found in loop regions. In contrast to glycine, proline has fixed torsion angles which are very close to those of the extended beta-sheet. That is why proline is often found at the end of helices and function as helix breaker. Theoretically, the average phi and psi values for α-helices and β-sheets should be clustered around −57 to −47 and −80 to +150, respectively.

2.6.1.1 Utility of Ramachandran Plot

Torsion angles normally fall in disallowed region (of Ramachandran plot) of bad or wrong homology models (Fig. 2.7, left figure of plot I) . That is why the Ramachandran plot is widely used in assessing the quality of experimental protein structures or structures built using homology modeling. Generally, the low resolution structures tend to have bad clustering due to bad geometry (Fig. 2.7, left figure of plot I), whereas high resolution structures have better clustering (Fig. 2.7, right figure of plot I). In bad structures, there is a lack of clustering around the allowed regions and also exhibit many dots in the disallowed regions (Fig. 2.7, left figure of plot II). Better structures have better clustering in allowed regions and almost none in the disallowed regions (Fig. 2.7, right plot II). Thus torsion angles outside the allowed regions should be examined carefully to resolve the problems in structures of protein.

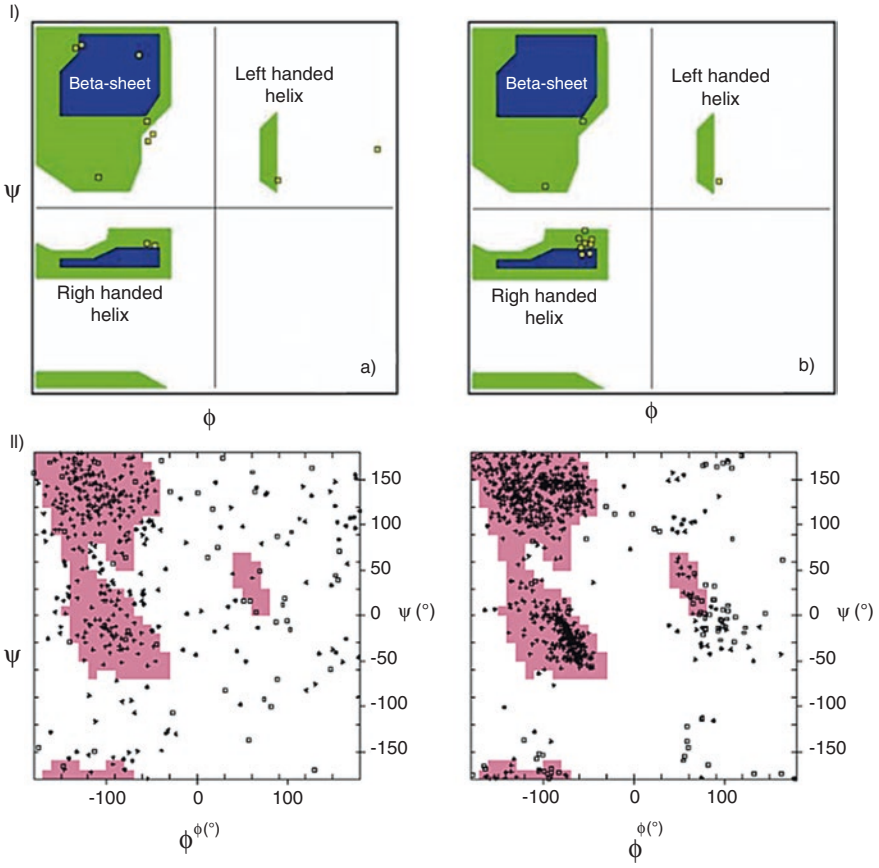

Fig. 2.7 Comparison of Ramachandran Plot. Plot I: Ramachandran plot of 11 amino acid residue protein (*left*). Mutation of one amino acid, tyrosine to serine, reduced the energetic state and stabilized the secondary structure to a-helix (*right*) (Reproduced from Akhmaloka et al. 2008). Plot II: Ramachandran plot of a protein structure obtained computationally; unrefined (*left*) and refined (*right*) (open access, Kleywegt and Jones 1998)

2.6.2 *Lattice Theory of Protein Folding*

In the absence of the availability of better sampling techniques, study of protein folding or identification of conformation can be done by using following approaches: (a) Empirical approaches, biophysical techniques such as X-ray, CD (Circular Dichroism), NMR (Nuclear Magnetic Resonance), SAXS (Small Angle X-ray Scattering), (b) Semi-empirical approaches, using bioinformatics, (c) computational approaches, such as MD and Monte Carlo simulation, and (d) simplification of conformational spaces.

Lattice theory is a simplified computer model to investigate and understand protein folding, which utilizes the concepts for simplifying conformational spaces. This is achieved by use of a 2D system as a model for the molecules in 3D space. In this model each residue is modeled as a bead or point and placed at the vertices of the lattice. To ensure the connectivity each residue on the backbone is placed on the same lattice vertex. Protein is considered as a linear chain of n amino acids. Each amino acid is either a polar (P) or non polar (H, Fig. 2.8) (Lau and Dill 1989). Discrete conformational space is using the lattice backbone. One important and valid assumption is that contact between HH has a free energy less than zero, and all the other combination has a free energy equal to zero.

2.6.3 *Energy Landscape Theory of Protein Folding*

The protein folding process sees the protein in terms of an energy landscape as depicted in Fig. 2.9 (Hartl et al. 2011). Energy landscape is a mapping of all possible conformations of protein folding comprising kinetic and thermodynamic intermediates. The key concept in this is folding funnel hypothesis (Fig. 2.9). Wolynes and colleagues (1991) proposed that the overall energy landscape of protein folding

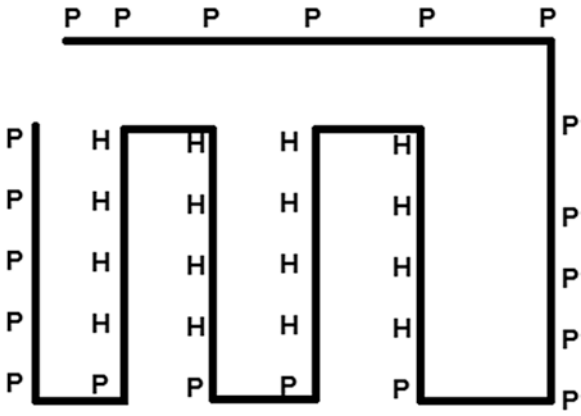

Fig. 2.8 2D representation of a model protein according to lattice model (Reproduced with permission, Lau and Dill 1989)

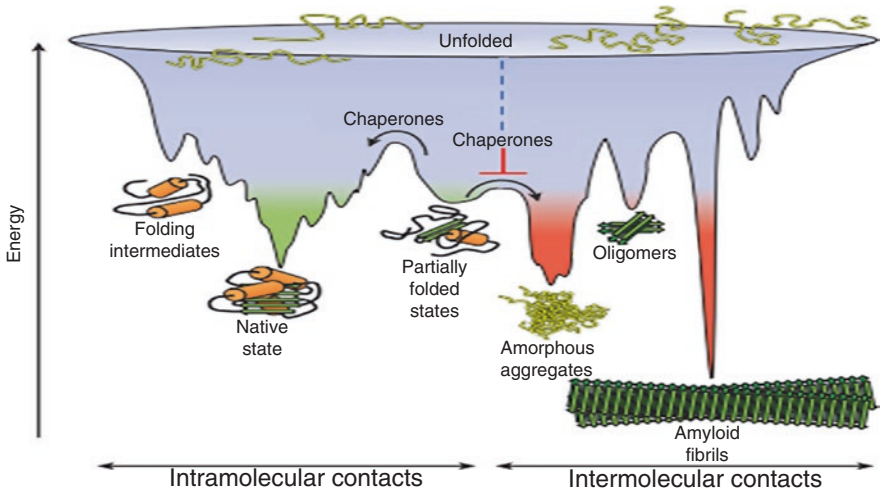

Fig. 2.9 Competing reactions of protein and aggregation. Scheme of the funnel-shaped free-energy surface that proteins explore as they move towards the native state (*green*) by forming intramolecular contacts. The ruggedness of the free-energy landscape results in the accumulation of kinetically trapped conformations that need to traverse free-energy barriers to reach a favorable downhill path. *In vivo*, these steps may be accelerated by chaperones. When several molecules fold simultaneously in the same compartments, the free-energy surface of folding may overlap with that of intermolecular aggregation, resulting in the formation of amorphous aggregates, toxic oligomers or ordered amyloid fibrils (*red*). Fibrillar aggregation typically occurs by nucleation-dependent polymerization. It may initiate from intermediates populated during *de novo* folding or after destabilization of the native state (partially folded states and is normally prevented by molecular chaperones) (Reproduced with permission, Hartl et al. 2011)

should be a funnel shape. No matter where one starts, it will follow the gradient towards a lower energy state. In a downhill path protein gradually accumulates favorable interactions, rejecting unfavorable ones, consequently reducing the search for native state. Newly synthesized peptides from the ribosome are in a high state of energy due to the flexibility and lack of any intramolecular structure, and tend to migrate to a lower energy state through a series of intramolecular or intermolecular interactions. This decrease in energy can include formation of secondary structures (α-helices or β-sheets), formation of oligomers, or formation of amorphous aggregates or amyloid fibrils (Hartl et al. 2011). Studies of chain entropies showed compactness is followed by lesser available conformations indicating that protein-folding energy landscapes are funnel-shaped. This protein energy funnel contains many local energy minima which allow for rapid interconversion with only small changes to the energy levels (Bryngelson et al. 1995). Protein folding landscapes are narrower at the bottom, suggesting that there are few low-energy, native-like conformations and many more open folded structures.

2.7 Molten Globule: An Important Intermediate in Protein Folding

Globular proteins are defined on the basis of their intrinsic viscosity and shape is approximately spherical. Their structural stability is due to the fact that non-polar residues are inside and polar residues are outside, promoting hydrophobic interactions (Dill 1990). In all globular proteins, the polarity of the faces of secondary structure is complimentary to the overall conformation of the proteins implying an interdependence of secondary structure and globular structure. When protein fold, from unfolded state to folded state, they go through a pathway consists of various intermediates. Two distinct intermediates can be identified in the folding pathway—kinetic and thermodynamic intermediate. Molten globule is one of the thermodynamically stable intermediate states, which exist between native and unfolded states. The resistance of such states depends upon a fine balance of stabilizing interactions such that the free energy difference between the native and intermediate states is less than that between the native and unfolded states, and this state appears to be less cooperative (Fig. 2.10). Sometime, the term "molten globule" is used for the state with similar but not identical, which accumulate during the folding process. These states may or may not have properties similar to molten globule.

Much of the detailed characterization of molten globule state is done under mild denaturation conditions where a protein is shown to have thermodynamically stable conformational state, between native and unfolded state. These states are found under acidic, basic, removal of cofactors, high temperature, and intermediate

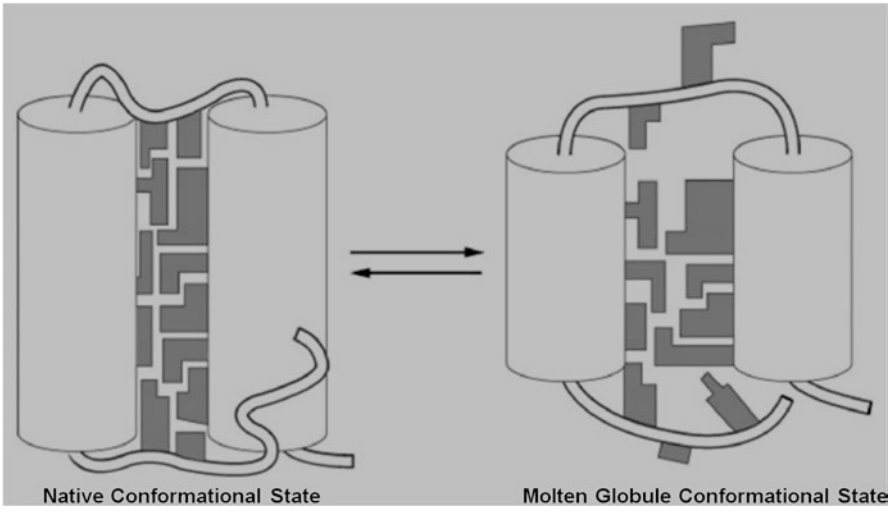

Fig. 2.10 Schematic representation of native and molten globule state (Reproduced from http://what-when-how.com/molecular-biology/molten-globule-molecular-biology/)

concentrations of urea or guanidine hydrochloride. The main characteristics of this state are:

1. Presence of secondary structure
2. Less stable tertiary structure
3. Globular shape
4. The tendency of Stable intermediates to aggregate,
5. Thermodynamic behavior as if "molten" states

One of the most important questions about the molten globule state is whether it is a specific thermodynamic state of protein molecule or whether it is similar to a slightly disordered native state or a squeezed coil. The existence of a thermodynamic state is based on the criteria that there should be a phase transition between two given states. So this first order transition is coupled with the drastic change of at least one of the thermodynamical parameters such as enthalpy.

As mentioned above, an intermediate is distinguished during folding, which is between native and unfolded state, and termed as "molten globule" based on defined characteristics (as listed above). Initially the term was referred as a wet molten globule—a conformational state which admits solvent and lacks close packing, while retaining native like secondary structure (Ptitsyn et al. 1968). However, the degree of side chain packing in molten globule is less clear. Protein denaturation is shown to be a first order phase transition (Privalov 1979), and Uversky and Ptitsyn (1994a, 1994b, 1996; Ptitsyn 1995) showed that the molten globule is separated by intermolecular first-order phase transitions from the native and unfolded states and therefore is a specific thermodynamic state of the protein molecules.

Recent observation indicated the existence of a closed packed molten globule conformation, which is an extended form of the native protein and does not allow solvent to penetrate the core—called dry molten globule (DMG). So the characteristic difference between DMG and MG is that the water is squeezed out of the core in the former. Although water cannot penetrate DMG but urea can (Hua et al. 2008). The governing thermodynamical features of close packing are, (a) a favorable change in conformational entropy that drives the process, (b) a compensating unfavorable change in enthalpy, and (c) an overall change in free energy that is much smaller than the change in $-T\partial S$, where T is temperature and S is entropy.

Molten globules are believed to have dynamic hydrophobic cores as evidenced by a non cooperative unfolding pattern, loss of signal in the near-UV CD spectrum, the broadening of NMR signals of aromatic groups (Dolgikh et al. 1981; Ohgushi and Wada 1983; Kjaergaard et al. 2010), rapid backbone amide exchange with solvents and high affinity for hydrophobic dyes (Kumar et al. 2014). Recent experiments demonstrated specific side chain interactions in α-lactalbumin in the molten globule state (Mok et al. 2005).

Similar to phase diagram of fluids, protein folding can be also viewed as a N/U/MG phase diagram (Fig. 2.11; Pande and Rokhsar 1998). Each state corresponds to a thermodynamic phase and the transition between two distinct phases corresponds to the exchange of stability following first order cooperative transition. Like the vapor-liquid, U-MG phase boundary appears to end at a critical point,

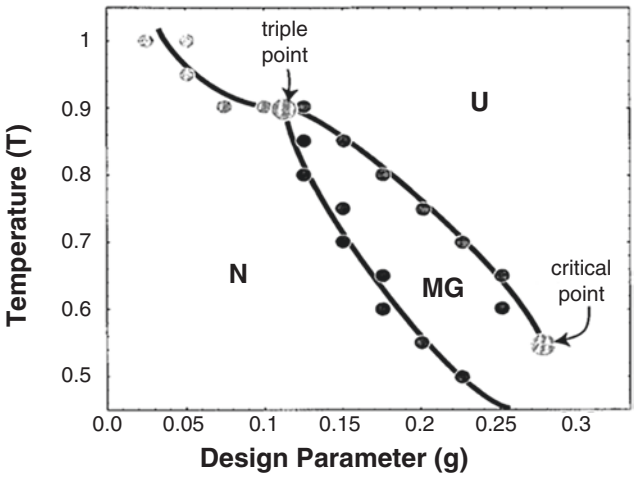

Fig. 2.11 Phase diagram of protein folding (Reproduced with permission, Pande and Rokhsar 1998; Copyright (1998) National Academy of Sciences, USA)

beyond which native interactions are destabilized (Uversky and Ptitsyn 1996; Schulman and Kim 1996).

The most interesting aspect of molten globule is its flexibility. Due to inherent flexibility this state has implications both for biological function and material science. The molten globule state has been confirmed by a number of experiments that the molten globule may be present in the cell and taking part in a number of cellular processes (Chap. 4). They have a potential role in genetic diseases (Chap. 4). Further research on molten globule will help to narrow the gap between protein physics and biology.

2.8 Concluding Remarks

Either by self assembly or with the help of chaperone proteins can acquire a well defined secondary and tertiary structures, involving two states (native and unfolded). The idea of the existence of only two states was challenged through demonstration of the existence of an intermediate state, molten globule. Existence of molten globule gave a whole new dimension to the five decade old problem of protein folding (PFP). PFP has evolved since its inception. Recent developments include statistical model and phase transition model, shifting the focus from a microscopic search to step-by step milestones on the pathways—unfolded state, transition state, native state and intermediate states. Another change in the approach has been the consideration of each state as an ensemble of conformations, rather than a single conformational state. These developments bring theory and experiments closer. However, the search of the generalized folding mechanism is not over yet. Probably we would

not be able to have a generalized mechanism due to the nature of the forces involved in a given folding process. We may have a better understanding and prediction of the model of experimentally established conformational states, which will help us to design better drugs against several diseases, especially those associated with folding problems.

References

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230
Akhmaloka, Susilowati PE, Subandi et al (2008) Mutation at tyrosine in AMLRY (GILRY like) motif of yeast eRF1 on nonsense codons suppression and binding affinity to eRF3. Int J Biol Sci 4:87–95
Baldwin R, Rose G (1999) Is protein folding hierarchic? II. Folding intermediates and transition states. Trends Biochem Sci 24:77–83
Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. Science 309:1868–1871
Bryngelson JD, Onuchic JN, Socci ND et al (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21:167–195
Camacho CJ, Thirumalai D (1993) Minimum energy compact structures of random sequences of heteropolymers. Phys Rev Lett 71:2505–2508
Chan HS, Shimizu S, Kaya H (2004) Cooperativity principles in protein folding. Methods Enzymol 380:350–379
Chou PY, Fasman GD (1978) Empirical predictions of protein conformation. Annu Rev Biochem 47:251–276
Daggett V (2002) Molecular dynamics simulations of the protein unfolding/folding reaction. Acc Chem Res 35:422–429
Daggett V, Fersht AR (2003) Is there a unifying mechanism for protein folding? Trends in Biochemical Sciences 28:18–25
DeBartolo J, Colubri A, Jha AK et al (2009) Mimicking the folding pathway to improve homology-free protein structure prediction. PNAS 106:3734–3739
Deleage G, Roux B (1987) An algorithm for protein secondary structure prediction based on class prediction. Protein Eng 1:289–294
de Prat GG, Ruiz-Sanz J, Neira JL et al (1995) Folding of a nascent polypeptide chain in vitro: cooperative formation of structure in a protein module. PNAS 92:3683–3686
Dill KA (1985) Theory for the folding and stability of globular proteins. Biochemistry 24:1501–1519
Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155
Dobson CM (2001) The structural basis of protein folding and its link with human disease. Philos Trans R Soc Lond B Biol Sci 356:133–145
Dolgikh DA, Gilmanshin RI, Brazhnikov EV et al (1981) Alpha-lactalbumin: compact state with fluctuating tertiary structure? FEBS Lett 136:311–315
Du R, Pande VS, Grosberg AU et al (1997) On the transition coordinate for protein folding. J Chem Phys 108:333–350
Epand RM, Scheraga HA (1968) The influence of long range interactions on the structure of myoglobin. Biochemistry 7:2864–2872
Faisca PFN, Gama MMT, Ball RC (2004) Folding and form: insight from lattice simulation. Phys Rev E 69:051917
Faisca PF, Plaxco KW (2006) Cooperativity and the origins of rapid, single-exponential kinetics in protein folding. Protein Sci 15:1608–1618
Flory PJ, Volkenstein M (1969) Statistical mechanics of chain molecules. Biopolymers 8:699–700

Fleming PJ, Rose GD (2005) Conformational properties of unfolded proteins. In: Kiefhaber T, Buchner J (eds) Protein folding handbook, vol 2. Wiley-VCH, Weinhein, pp 710–736
Frauenfelder H, Sligar SG, Wolynes PG (1991) Energy landscape and motion of proteins. Science 254:1598–1603
Frauenfelder H, Wolynes PG (1994) Biomolecules: where the physics of complexity and simplicity meet. Phys Today 47(2):58–64
Gorka A (2011) Prediction and application of intrinsically disordered regions in practice. http://en.bioinformatyk.eu/contest-articles/prediction-and-application-of-intrinsically-disordered-regions-in-practice.html.
Guo Z, Thirumalai D, Honeycutt J (1992) Folding kinetics of proteins—a model study. J Chem Phys 97:525–535
Guo ZY, Thirumalai D (1995) Kinetics of protein-folding-nucleation mechanism, time scales, and pathways. Biopolymers 39:83–102
Hao MH, Scheraga HA (1994) Monte-Carlo simulation of a first order transition for protein folding. J Phys Chem 98:4940–4948
Hao MH, Scheraga H (1998) Molecular mechanisms for cooperative folding of proteins. J Mol Biol 277:973–983
Hartl FU, Bracher A, Hartl MH (2011) Molecular chaperones in protein folding and proteostasis. Nature 475:324–332
Hua L, Zhou R, Thirumalai D et al (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. PNAS 105:16928–16933
Kleywegt GJ, Jones TA (1998) Databases in protein crystallography. Acta Cryst D54:1119–1131
Kjaergaard M, Teilum K, Poulsen FM (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. PNAS 107:12535–12540
King RD, Stenberg MJ (1996) Identification and application of the concepts important for accurate and reliable protein structure prediction. Protein Sci 5:2298–2310
Klimov DK, Thirumalai D (1998) Lattice models for proteins reveal multiple folding nuclei for nucleation-collapse mechanism. J Mol Biol 282:471–492
Kumar R, Kukreja RV, Cai S et al (2014) Differential role of molten globule and protein folding in distinguishing unique features of botulinum neurotoxin. Biochim Biophys Acta 1844:1145–1152
Lau KF, Dill KA (1989) A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules 22:3986–3997
Levin JM (1997) Exploring the limits of nearest neighbor secondary structure prediction. Protein Eng 10:771–776
Li L, Mirny LA, Shakhnovich EI (2000) Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. Nat Struct Biol 7:336–342
Lin MM, Zewail AH (2012) Hydrophobic forces and length limit of foldable protein domains. PNAS 109:9851–9856
Matveev M (2010) Native aggregation as a cause of origin of temporary cellular structures needed for all form of cellular activity, signaling and transformation. Theor Biol Med Model 7:1–22
Mok KH, Nagashima T, Day IJ et al (2005) Multiple subsets of side chain packing in partially folded states of alpha-lactalbumin. PNAS 102:8889–8904
Ohgushi M, Wada A (1983) Molten-globule state—a compact form of globular proteins with mobile side chains. FEBS Lett 164:21–24
Pande VS, Rokhsar DS (1998) Is the molten globule a third phase of proteins? Biochemistry 95:1490–1494
Peng Q, Fang J, Wang M et al (2011) Kinetic partitioning mechanism governs the folding of the third FnIII domain of tenascin-c: evidence at the single-molecule level. J Mol Biol 412:698–709
Prieto L, Rey A (2007) Influence of the chain stiffness on the thermodynamics of a Gō-type model for protein folding. J Chem Phys 127:175101
Privalov PL (1979) Stability of proteins: small globular proteins. Adv Protein Chem 33:167–241
Privalov PL (1989) Thermodynamic problem of protein structure. Annu Rev Biophys Biophys Chem 18:1–47

Ptitsyn OB, Kron A, Eizner YY (1968) Simple statistical theory of self-organization of protein molecules. J Polym Sci C 16:3509–3516

Ptitsyn OB (1995) Structures of folding intermediates. Curr Opin Struct Biol 5:74–78

Radford SE (2000) Protein folding: progress made and promises ahead. Trends Biochem Sci 25:611–618

Ramachandran GN, Ramakrishnana C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99

Ramanathan S, Shakhnovich E (1994) Statistical mechanics of proteins with "evolutionary selected" sequences. Phys Rev E 50:1301–1312

Rost B, Sander C, Schneider R (1994) PHD: an automatic mail server for protein secondary structure prediction. Comput Appl Biosci 10:53–60

Schulman BA, Kim PS (1996) Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology. Nat Struct Biol 3:682–687

Shakhnovich E, Finkelstein A (1989) Theory of cooperative transitions in protein molecules, I. Why denaturation of globular protein is a first order phase transition. Biopolymers 28:1667–1681

Shakhnovich EI (1998) Protein design: a perspective from simple tractable models. Fold Des 3:R45–R58

Simons KT, Bonneau R, Ruczinski I et al (1999) Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins Suppl 3:171–176

Sfatos CD, Gutin AM, Shakhnovich EI (1993) Phase diagram of random copolymers. Phys Rev E 48:465–475

Sosnick TR, Mayne L, Hiller R et al (1994) The barriers in protein folding. Nat Struct Mol Biol 1:149–156

Stigler J, Zeigler F, Gieseke A et al (2011) The complex folding network of single calmodulin molecules. Science 334:512–516

Tanford C (1968) Protein denaturation. Adv Protein Chem 23:121–282

Thirumalai D (1995) From minimal models to real proteins: time scales for protein folding kinetics. J Phys I 5:1457–1467

Uversky VN, Ptitsyn OB (1994a) "Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of β-lactamase at low temperature. Biochemistry 33:2782–2791

Uversky VN, Ptitsyn OB (1994b) The molten globule is a third thermodynamical state of protein molecules. FEBS Lett. 341: 15–18

Uversky VN, Ptitsyn OB (1996) Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. Fold Des 1:117–122

Yang JS, Chen WW, Skolnick J et al (2007) All-atom *ab initio* folding of a diverse set of proteins. Structure 15:53–63

Chapter 3
Relevance of Intrinsic Disorder in Protein Structure and Function

Thomas M. Feltrup, Raj Kumar, and Bal Ram Singh

Keywords Intrinsically disordered proteins • IDPs • Molten globule • Botulinum neurotoxin

3.1 Definition of Intrinsically Disordered Proteins (IDPs)

A traditional view of protein folding requires a well-defined secondary structure (α-helices and β-sheets) forming a compact, globular tertiary structure; this globular structure determines what the protein's target and function will be (Wright and Dyson 1999; Dunker et al. 2002; Uversky 2002a, b). It has been generally considered that proteins would tend to avoid any unfolding or disorder in the active state since this would result in an increase in aggregation or misfolding with a loss of activity (Braselmann et al. 2013). This "lock-and-key" classification works for many proteins. However, functional proteins having some disordered regions or completely disordered has become a recently accepted idea in proteomics (Tompa 2011; Uversky 2013b). The significant number of functional proteins with at least some disorder has changed the way protein structure is classified. An 'ordered' protein no longer consists only of proteins with a well-defined globular structure, while a 'disordered' protein does not necessarily imply an inactive protein. With disordered states now shown to be active, the differentiation in secondary and tertiary structures between disordered proteins and ordered proteins was described over a decade ago (Fig. 3.1; Dunker et al. 2001). This is a modification of the "structure-function paradigm" which requires a defined 3D structure before a function is possible. The number of proteins with at least some disorder continues to climb as better techniques and predictors are developed (Dyson and Wright 2005; Tompa 2012).

R. Kumar, B.R. Singh, *Protein Toxins in Modeling Biochemistry*, SpringerBriefs in Biochemistry and Molecular Biology, DOI 10.1007/978-3-319-43540-4_3

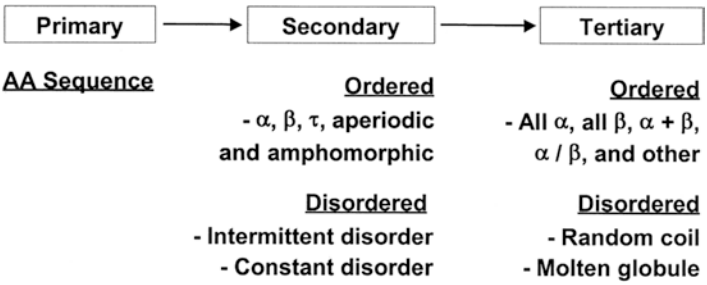

Fig. 3.1 Primary, secondary, and tertiary structure classification of ordered and disordered proteins. At the secondary structure level, disordered proteins can exist in either intermittent disorder, such as a disordered-to-ordered transition upon binding, or in constant disorder where the protein maintains disorder in both the resting and active state. At the tertiary level, disordered proteins can either maintain some loose, global structure, the molten globule state, or be completely disordered and exist in a random coil state (Reproduced with permission, Dunker et al. 2001)

Intrinsically disordered proteins (IDPs) are characterized in two major categories. The first category includes proteins/regions resembling random coils. This characterization can encompass both a global structure of the protein or a localized region of a protein. Proteins in this group have little to no secondary or tertiary structure in the disordered region, resembling a denatured state of an ordered protein (Uversky 2002a; Tompa 2005). These proteins, or regions, have no clear equilibrium state and maintain dynamic backbone torsion angles; their energy landscapes contain many local minima without a global minimum, unlike an ordered protein (Uversky 2013a). While these proteins rapidly interconvert between conformations there are some distinct characteristics for all IDPs (Tompa 2005). IDPs identified have a low sequence complexity with a distinct amino acid composition where the proteins have a higher disorder-promoting amino acid composition (A, R, G, Q, S, P, E, and K) while lacking order-promoting amino acids (W, C, F, I, Y, V, L, and N) (Fig. 3.2) (Dunker et al. 2001; Tompa 2005; Dyson and Wright 2005). To better maintain a flexible, disordered structure, the IDPs typically contain a high global net charge, typically negative, along with a low mean hydrophobicity (Hemmings et al. 1984; Tompa 2005). B-factors appear to also be correlated with the abundance of certain amino acids in IDPs. Amino acids with higher B-factors show a higher degree of electron density smearing in high-resolution X-ray crystallography and therefore a more flexible environment (Parthasarathy and Murthy 2000). As expected, amino acids with higher B-factors are more abundant in IDPs (B-factor increases left to right in Fig. 3.2).

A second group of IDPs can be classified as molten globules (including the pre-molten globule proteins). The molten globule state is a conformation composed of structural elements somewhere between the ordered and denatured states of the protein. The secondary structure of the molten globule state has been characterized as

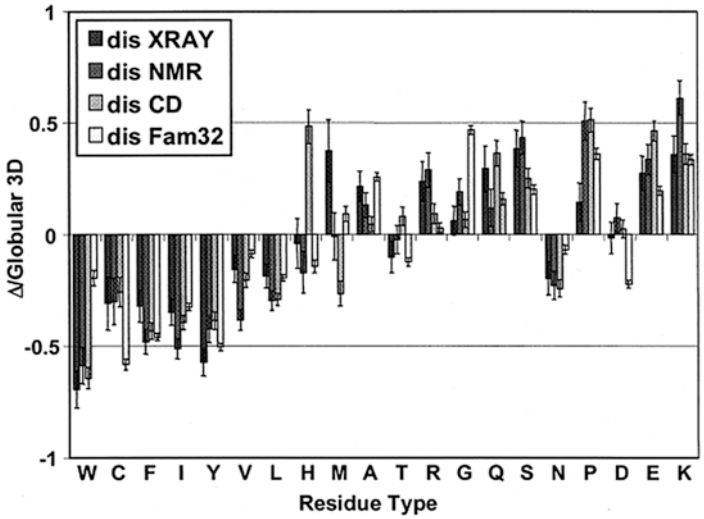

Fig. 3.2 Amino acid composition profile. The prevalence of each amino acid in disordered regions/proteins compared to ordered proteins calculated based on proteins characterized by X-ray crystallography, NMR, CD, or homology based on an initial set of 32 proteins. A *negative peak* indicates depletion of that amino acid in disordered regions/proteins while a *positive peak* indicates enrichment in that amino acid in disordered regions/proteins. Residues are arranged in increasing residue flexibility from *left* to *right* based on B-factor estimates, also known as Debye-Waller temperature factors (Reproduced with permission, Dunker et al. 2001)

near "native-like" in which the secondary structure based on far-UV CD, NMR, and IR spectroscopy is identical or nearly identical to the folded state (Dolgikh et al. 1981; Ohgushi and Wada 1983; Ptitsyn 1987). The distinguishing characteristic of the molten globule state is a loosened tertiary structure; this loosened structure is easily visualized using near-UV CD (shown by a decrease in signal), the increased exposure of the hydrophobic core (measured using 1-anilino-8-napthalene sulfonate (ANS) binding), (Dolgikh et al. 1981; Ohgushi and Wada 1983; Dunker et al. 2001; Cai and Singh 2001) and deuterium exchange (Dolgikh et al. 1981, 1984, 1985). While not as flexible as the random coil proteins, molten globule proteins do show a significant amount of disorder in their domains with the pliable nature of the tertiary structure allowing for many of the same advantages.

As of 2003, only 30 % of the proteins listed in PDB were fully ordered with no disordered regions (Obradovic et al. 2003). With the rapidly expanding library of disordered proteins, the DisProt database was created to provide structural and functional information on both types of IDPs—partially disordered proteins (molten globules and pre-molten globules), and fully disordered proteins (Vucetic et al. 2005). With disordered proteins now more centrally located in a database, it has become more convenient to compare the frequency of disordered proteins across different cellular systems all the way up to different kingdoms of life.

3.2 Experimental Tools to Identify Intrinsically Disordered Proteins

3.2.1 X-ray Crystallography

X-ray crystallography is a powerful tool when determining the structure of a protein crystal. This technique can provide a detailed map at the atomic scale with the locations of protein secondary and tertiary structure. While a powerful and popular technique, X-ray crystallography has several important limitations when dealing with IDPs. Proteins in a crystal state are required for characterization using X-ray crystallography; crystallography of disordered regions of the protein, whether the disorder is local (such as a loop or small region) or global, becomes very difficult, if not impossible (Dunker et al. 2001).

With these limitations, structural details cannot be resolved for these regions or proteins; however, locating the potential disorder in proteins is possible by examining the residues missing in the electron density maps, such as in the case of the ammonia channel, glycerol uptake facilitator protein, and calcineurin (Fig. 3.3)

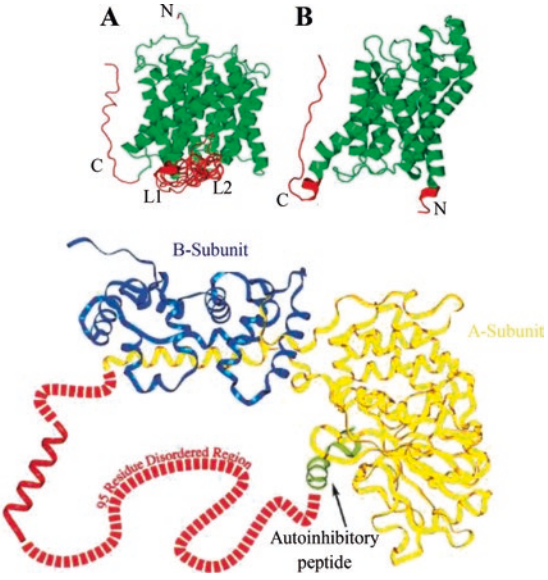

Fig. 3.3 Protein structures containing regions of disorder. Regions colored in *red* are intrinsically disordered. (**a**) The ammonia channel (PDB ID: 2NMR) has disordered regions at N-terminal Ala1, 182–194, 302–310, and 387–406. (**b**) Glycerol uptake facilitator protein (PDB ID: 1FX8) shows disorder at the N-terminal (1–5) and the C-terminal (260–281). (**c**) Calcineurin is composed of two subunits with a large region of disorder. The A subunit (*yellow*) contains a phosphatase domain and binds to the B subunit via a helix extension. A 95 residue region of the protein connecting the autoinhibitory peptide to the helix extension of the A subunit was identified by a lack of electron density in the crystal structure and therefore is unobserved (Reproduced with permission, Xue et al. 2009 and Dunker et al. 2001)

(Romero et al. 2000; Dunker et al. 2001; Xue et al. 2009). As research on disordered regions and proteins continues to grow and the importance of this disorder becomes better understood, the limitations of X-ray crystallography have become more profound and must be used along with other tools to get a complete picture.

3.2.2 Small Angle X-ray Scattering (SAXS)

Small angle X-ray scattering (SAXS) spectroscopy allows for the determination of the overall size and shape of a protein molecule in solution (Eliezer et al. 1993). SAXS is a time-resolved technique which can characterize the kinetics of processes and reactions. Other useful information gathered from data is the protein molecular weight, excluded particle volume, maximum dimension (D_{max}), and radius of gyration (R_g) (Kikhney and Svergun 2015). Proteins can be qualitatively identified by the data as globular or disordered by a Kratky Plot (Fig. 3.4); from the data, a low resolution model can be reconstructed with further validation from higher resolution techniques such as X-ray crystallography and NMR spectroscopy (Kikhney and Svergun 2015).

While this technique can only give the overall size and shape of the protein and not information at the atomic level like X-ray crystallography, when dealing with proteins with disorder this technique provides some structural information which is not possible with X-ray crystallography.

3.2.3 Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) spectroscopy is an important and versatile tool for studying protein structure. CD spectroscopy can be separated into far-UV and near-UV; far-UV CD can determine the overall secondary structure of a protein in solution

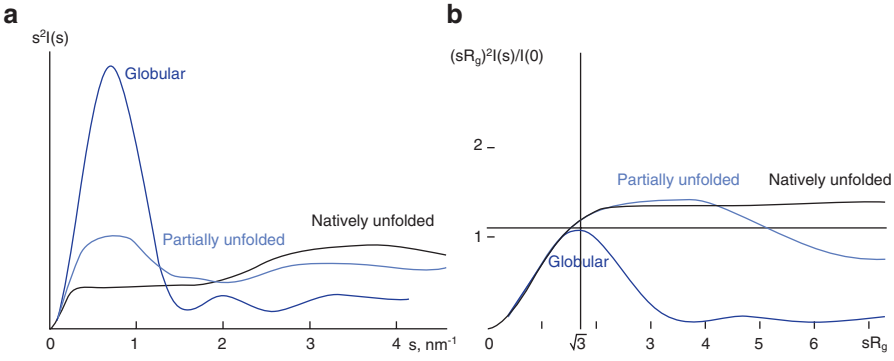

Fig. 3.4 Simulated data from three 60 kDa proteins. Folding includes globular (*dark blue*), 50 % unfolded (*light blue*), and fully disordered (*gray*). (**a**) Kratky plot of $s^2I(s)$ vs. s. (**b**) A normalized Kratky plot $(sR_g)^2I(s)/I(0)$ vs. sR_g (Reproduced with permission, Kikhney and Svergun 2015)

while near-UV CD can help in determining the presence of tertiary structure. The diminished secondary structure can be monitored by observing the signal at about 210 and 240 nm (Romero et al. 2000). CD spectroscopy is only semi-quantitative and unlike X-ray crystallography, it is not able to give structural information at the residue level (Dunker et al. 2001); however, CD spectroscopy does allow monitoring of protein structures in solution as well as obtaining information for disordered proteins which X-ray crystallography is unable to.

3.2.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy is a very sensitive technique which can provide dynamic information about proteins on the order of 10^{-12}–10^{5} s and has been used in proteins as large as the 900 kDa GroEL complex (Fiaux et al. 2002; Boehr et al. 2006). NMR provides the benefit of determination of structures in a solution state while also providing data at the residue level (Ishima and Torchia 2000; Dunker et al. 2001); ^{15}N-^{1}H heteronuclear NOE measurements are able to detect the difference between ordered residues and disordered residues based on the value of the signal (Evans 1995; Ishima and Torchia 2000; Bracken 2001). An example of IDP and ordered protein ^{1}H-^{15}N HSQC spectra can be found in Fig. 3.5. These benefits do have limitations as NMR requires a relatively high protein concentration in which aggregation does not occur; NMR analysis can also be difficult with disordered proteins due

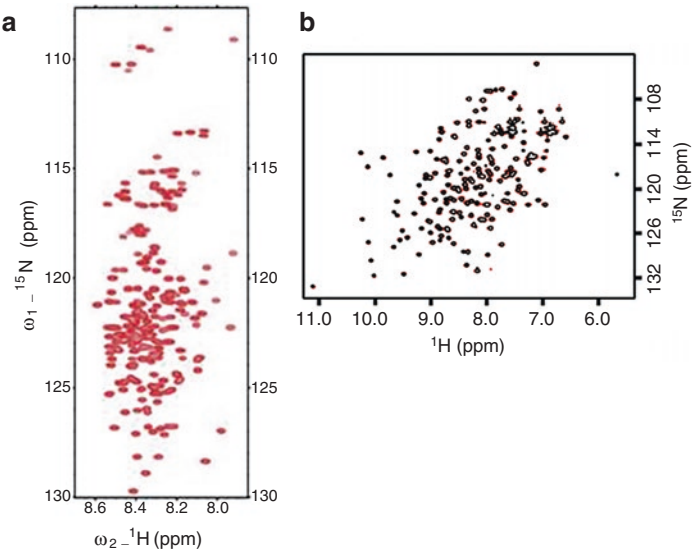

Fig. 3.5 $^{1}H^{N}$-^{15}N HSQC spectra of (**a**) the intrinsically disordered protein BASP1 and (**b**) the ordered protein AlgH. $^{1}H^{N}$-^{15}N HSQC spectra of IDPs are characterized by a narrower spectral range than ordered proteins (Reproduced with permission, (**a**) Konrat 2014; (**b**) Urbauer et al. 2015)

to resonance broadening or lack of chemical shift dispersion (Dunker et al. 2001; Fontana et al. 2004). Despite these limitations, NMR spectroscopy has been a vital tool in the characterization of disordered proteins.

3.2.5 *Protease Digestion*

The general flexibility of a specific protein can be monitored using a protease to digest the sample (Fontana et al. 1993, 1997). Proteolysis of the target of interest can occur if the protease can bind to the target (typically about a ten residue unfolded region) (Hubbard et al. 1994) and adapt to the stereochemistry (Dunker et al. 2001). Rigid peptide structures are not easily bound by proteases and therefore not cleaved; more flexible structures on the other hand are more readily bound and cleaved by proteases. The amount of cleaved peptides present in a digested sample is related to the flexibility of the protein target. Figure 3.6 shows an example of a protein digested with trypsin; regions are marked with white lines to show all potential cleavage sites and sites not cleaved suggest a more rigid region in the protein. While this

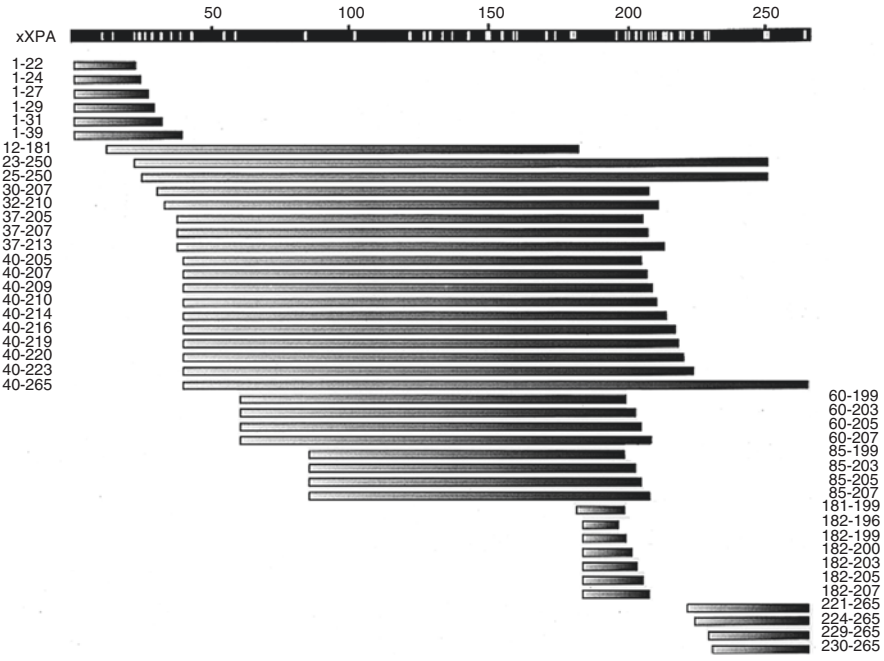

Fig. 3.6 All potential trypsin cleavage sites of xXPA are identified by *white lines* in the *black line* representing the full-length xXPA. Fragments were identified by ESI-FTICR mass spectrometry; amino acid positions for each fragment are listed (Reproduced with permission, Iakoucheva et al. 2002)

technique, like many of the other techniques, does not give specific information at the atomic level, it can be combined with techniques such as mass spectrometry to help determine the regions of higher disorder (Iakoucheva et al. 2001).

3.2.6 Hydrodynamic Radius

Proteins with abnormally high radii compared to their molecular weight typically show disorder to some degree. While chemically denatured proteins exhibit predictable behavior based on random coils, IDPs can possess non-random structure (Marsh and Forman-Kay 2010). The hydrodynamic radius can be calculated using equation 1 and measured using techniques such as SAXS or gel filtration:

$$\mathrm{R_h} = \frac{k_b T}{6\pi D} \tag{3.1}$$

The R_h has been used to identify disorder based on an abnormally high hydrodynamic radius (Hernàndez et al. 1986; Weinreb et al. 1996). Marsh and Forman-Kay plotted protein length vs. R_h values for folded, chemically denatured, and intrinsically disordered proteins (Fig. 3.7) and then proposed a relationship in the following equations in which N is the number of residues:

$$R_h^{folded} = 4.92N^{0.285} \tag{3.2}$$

$$R_h^{denatured} = 2.93N^{0.549} \tag{3.3}$$

$$R_h^{IDP} = 2.49N^{0.509} \tag{3.4}$$

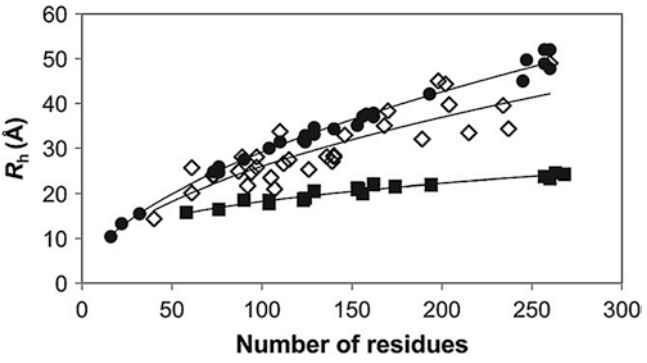

Fig. 3.7 Protein size vs. R_h values of 20 folded proteins (*solid squares*), 27 chemically denatured proteins (*solid circles*), and 32 IDPs (*open diamonds*) (Reproduced with permission, Marsh and Forman-Kay 2010)

If a larger than anticipated radius is observed, other techniques like the ones listed above are used in tandem to better understand the amount and location of disorder in the protein (Marsh and Forman-Kay 2010).

3.3 Protein Structural Disorder Predictors and Disorder Database

Since direct observation and characterization of disordered regions can be difficult with many of the current experimental techniques using a prediction based technique can be highly useful. There are a number of available prediction tools to identify ordered/disordered regions in protein sequences which are briefly described below.

3.3.1 Prediction of Natural Disordered Residues (PONDR)

The prediction of natural disordered residues, or PONDR, was initiated after studies on calcineurin (CaN) which showed long regions of disorder. From this result, Romero et al. (1997, 2001) designed a series of rules to better predict disorder in proteins based on the sequence. The low aromatic amino acid content was a common occurrence in disordered regions and was the first rule adopted into the PONDR prediction. If a contiguous region of more than 80 amino acids in a protein have an average aromatic amino acid content of less than 6.5 % then that region is predicted to have a long disordered region (LDR), this is the rule-based predictor portion of PONDR.

A neural network predictor was developed for varying lengths of disordered regions. After compiling a sufficiently large set of proteins with at least a disordered region of seven residues based on the Protein Data Bank (PDB), the disordered regions were classified as short disordered regions (SDR, 7–21 residues), medium disordered regions (MDR, 22–44 residues), and long disordered regions (LDR, 45+ residues). Based on this set of disordered proteins, the additional defined characteristics are low amounts of both Cys and His and high amounts of Glu, Asp, Ser, and Lys. Low amounts of Cys prevent disulfide bridges to maintain more flexibility in the region; High amounts of charged residues promote charge imbalance and promote disorder in the region while Ser increases the solubility of the region.

The prediction accuracy of each of these neural network predictors (SDR, MDR, and LDR) had an approximately 70 % success rate when predictors were used on the same length class of protein; however, when predicting different length class of proteins, the prediction accuracy fell. Using this framework, Li et al. (1999) were able to improve predictions of disorder in proteins based on whether the disorder was located near the C-terminus, N-terminus, or located internally.

Each of the PONDR predictors can be accessed at: http://www.pondr.com/.

3.3.2 *FoldIndex*

FoldIndex is based on a predictor first described by Uversky et al. (2000). The equation proposed included the mean net charge (|‹R›|) and mean hydrophobicity (‹H›) based on the Kyte/Doolittle scale. Prilusky et al. (2005) rearranged Uversky's equation to give the FoldIndex equation:

$$I_F^{KD} = 2.785\,H - |R| - 1.151.$$

Using this equation (the known boundary of disordered vs. ordered) and combining a boundary created using the logical limits of the charge-hydropathy space (|‹R›| = 1.125 – 1.125 ‹H›), an accessible area plot was created (Fig. 3.8).

A positive value represents a folded protein/domain where a negative value represents an unfolded protein/domain. This predictor was tested on 39 reported disordered proteins as well as 151 ordered proteins from the OCA Protein Data Bank. FoldIndex correctly predicted the unfolded structure of the 39 unfolded protein set 77 % of the time; this success rate was higher than PONDR, DISOPRED, and GlobPlot. When used to predict the 151 folded proteins, FoldIndex correctly predicted the proteins 88 % of the time. This prediction was not as accurate as the other methods used, however; DISOPRED (98 %), PONDR (93 %), and GlobPlot (99 %) all outperformed FoldIndex in correctly predicting the folded proteins. A sample of FoldIndex can be seen in Fig. 3.9.

The FoldIndex can be accessed at: http://bip.weizmann.ac.il/fldbin/findex.

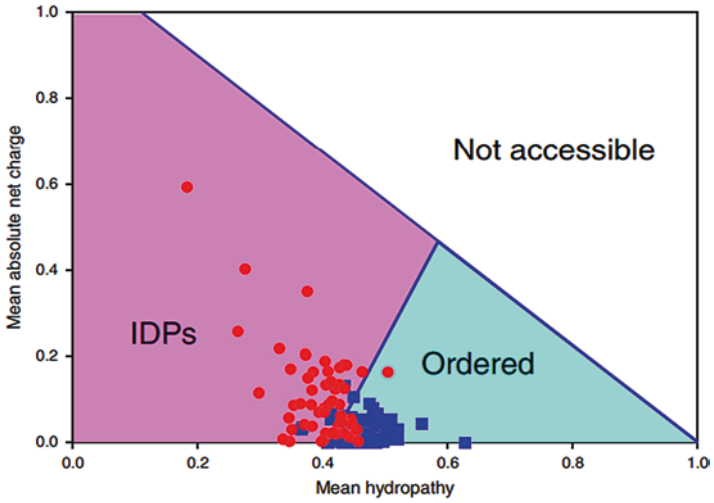

Fig. 3.8 Charge-hydropathy plot for ordered proteins (*blue squares*) and IDPs (*red circles*). Available space for ordered proteins is shown in *cyan*; available space for IDPs are in *pink* (Reproduced with permission, Uversky 2013b)

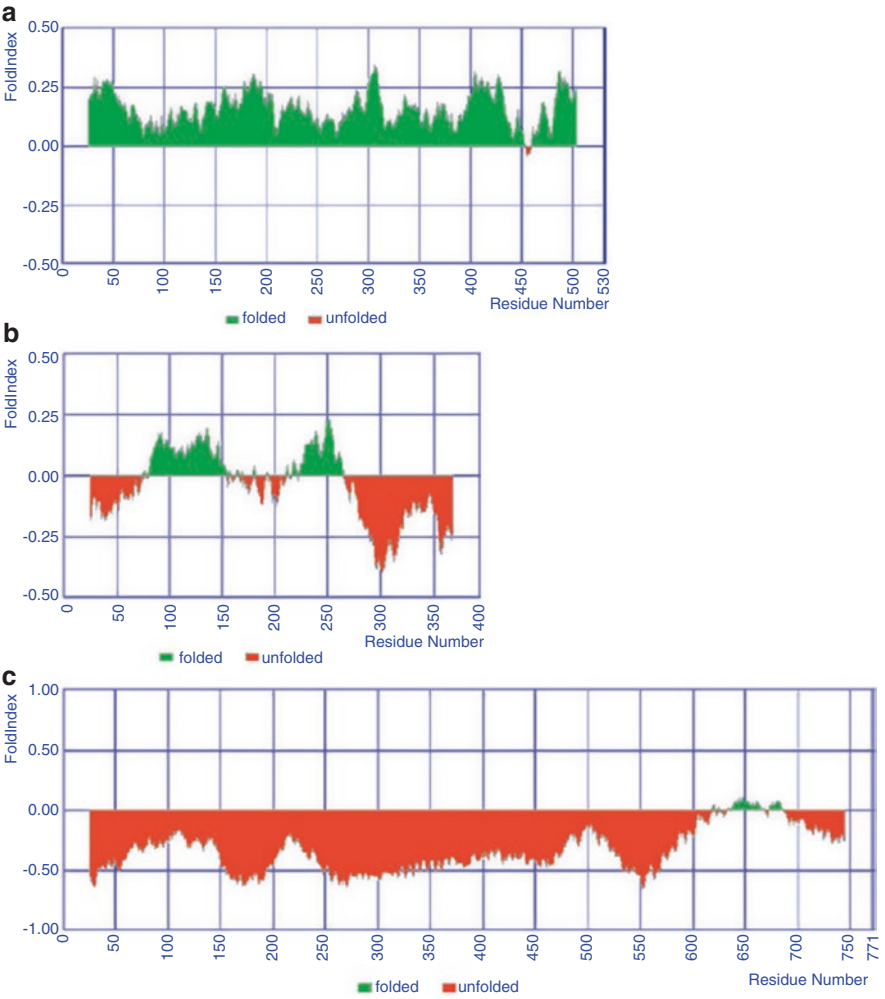

Fig. 3.9 FoldIndex plotted for (**a**) Cat-muscle (M1) pyruvate kinase which shows nearly no disorder and is folded (PDB-ID 1PKM; Swiss-Prot: P11979), (**b**) Human p53 tumor suppressor protein has large regions of disorder with some folded regions (PDB-ID 1TSR; Swiss-Prot P04637), (**c**) Chicken gizzard caldesmon is unfolded (Swiss-Prot 12957) (Reproduced with permission, Prilusky et al. 2005)

3.3.3 DISOPRED2

The DISOPRED2 prediction method was designed by Ward et al. (2004) to analyze genomes for disordered proteins and then investigate which processes were most influenced by the presence of disordered proteins (Fig. 3.3). The accuracy of DISOPRED2 is largely due to the fact that it considers disorder promoting

sequences, such as Pro-X-Pro-X-Pro or Lys-X-X-Lys-X-Lys, rather than amino acid composition, hydrophobicity, or charges. The DISOPRED2 system was used to classify the proteome of representative genomes from eubacteria, archaea, and eukaryotes (Chen et al. 2006; Pancsa and Tompa 2012). The increased abundance of disorder in eukaryotic genomes was observed and discussed further in this paper.

3.3.4 DisEMBLE

The DisEMBLE prediction tool uses a combination of three main criteria to predict disorder within a protein developed by Linding et al. (2003). The first is whether the region is found in a loop or coil; these regions found in a loop or coil tend to be disordered. The second is based on hot loops which use the C_α temperature factors (B factors) that predict high levels of mobility in the loop. The final factor is the crystal structure located in PDB which has the restriction of having only published structures along with many crystals using truncated versions of proteins due to flexibility/disorder.

This composite prediction tool is able to predict >60 % of hot loops with a low error rate. The DisEMBLE predictor can be accessed at: http://dis.embl.de/.

3.3.5 GlobPlot

The GlobPlot disorder prediction is a propensity based tool. Linding et al. (2003) developed an algorithm based on propensities of amino acids, called the Russell/Linding propensities, to be in a globular or random coil/disordered state. The GlobPlot algorithm initially correctly predicted the presence of disorder in all 20 selected proteins from the review paper by and compared favorably to other predictors such as PONDR and SMART. The GlobPlot predictor can be accessed at: http://globplot.embl.de/.

3.3.6 DisProt Database

A centrally compiled database of proteins with complete disorder along with proteins with smaller sections of disorder was created in 2005 (Vucetic et al. 2005). Included proteins are mainly identified by X-ray crystallography where regions showing missing electron density along the backbone are further investigated (Vucetic et al. 2005). Most proteins classified in DisProt have been analyzed by multiple methods to confirm the disorder; these methods include NMR, CD spectroscopy, SAXS, protease digestion, and the hydrodynamic radii (Vucetic et al. 2005; Sickmeier et al. 2007).

The DisProt database works in tandem with other databases including PDB, SWISS-PROT, TrEMBL, GenBank, and PIR by providing links to proteins found in multiple databases (Vucetic et al. 2005). The description of the protein includes, among other items, a description of the function, map of the disordered regions, and analysis of the disordered regions (Sickmeier et al. 2007).

3.4 Classification and Examples of Intrinsically Disordered Proteins

The number of proteins classified as IDPs has expanded considerably over the last decade as more predictors have been developed. With more information being gathered, classifications of the functional, structural, and regulatory roles of IDPs can be made (Fig. 3.10). IDPs can be classified as falling into one of four general categories: entropic chains, protein modification, molecular assembly, and molecular recognition (Dunker et al. 2002). Further subdivisions can be created based on mechanism, structure and targets. A brief overview of the function of each of the four major classes is given below; a brief list of disordered proteins with their disordered regions is presented in Table 3.1. More extensive lists of IDPs along with their detection method and function have also been described (Dunker et al. 2002; Uversky 2002a, b; Sickmeier et al. 2007; Romero et al. 1998).

3.4.1 Entropic Chains

The functions of entropic chain IDPs rely directly on the flexible conformational state and perform functions unable to be performed by a folded state (Dunker et al. 2002; Tompa 2005). Entropic chains can function as springs, bristles and spacers, linkers, clocks, and possibly more (Dunker et al. 2002; Uversky 2015b). Some of the proteins falling into this category include titin (spring), K^+ channel ball and chain (clock), fd pIII protein (linker), and neurofilament H (bristle) (Dunker et al. 2002).

Entropic springs, such as titin, provide a structural support in which a PEVK-enriched disordered region can help restore overstretched muscle cells (Trombitás et al. 1998; Dunker et al. 2002). A flexible linker found in tripartite Tc toxin was shown to exist as an entropic spring in its native state and aid in the insertion of the toxin into the host membrane by contracting (Fig. 3.11) (Meusch et al. 2014).

Entropic clocks, such as the K^+ channel ball and chain, can use disorder to control entry to a channel in response to a stimulus like a voltage gradient (Antz et al. 1997; Dunker et al. 2001; Zandany et al. 2015).

Linkers are flexible regions of proteins, typically exposed, connecting domains. These linkers are susceptible to protease digestion, suggesting the exposed nature of

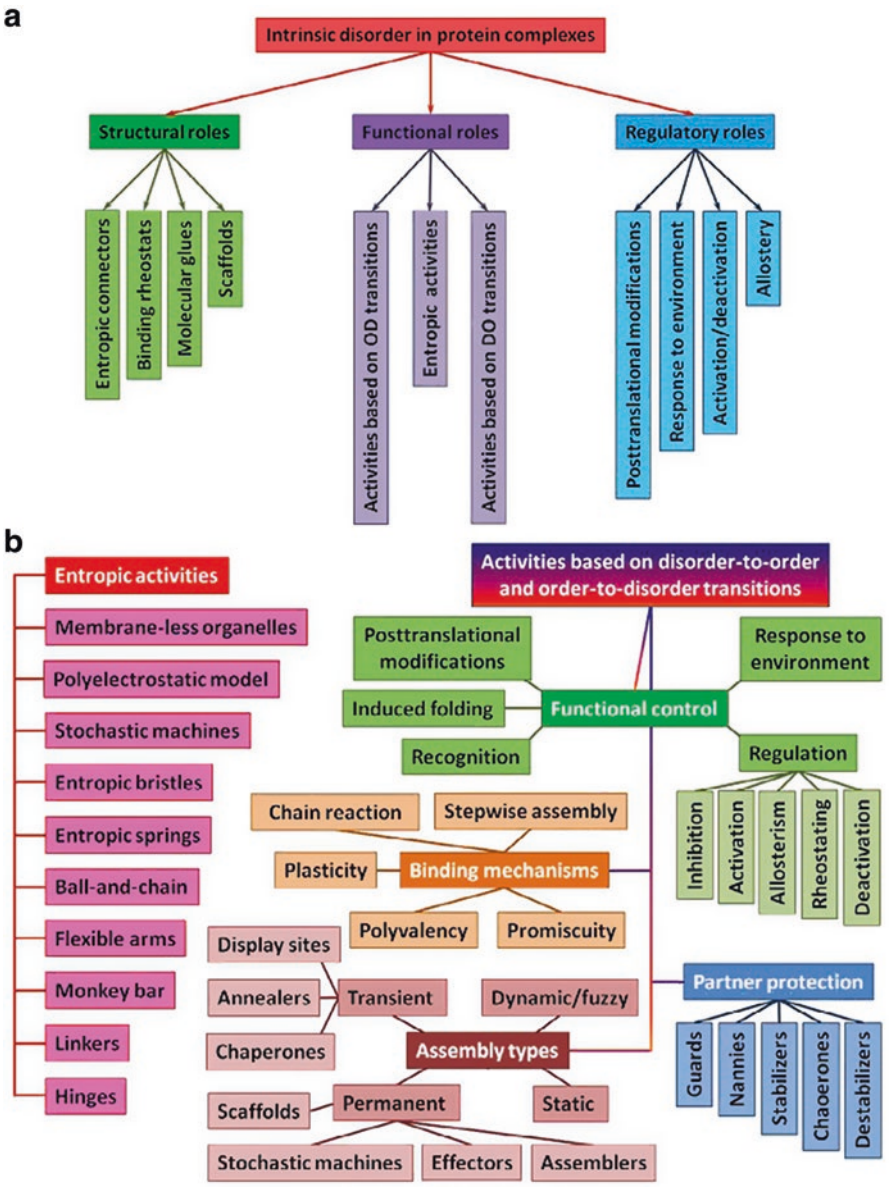

Fig. 3.10 (**a**) The general functional, structural, and regulatory roles of IDPs. (**b**) The entropic roles and activities of IDPs based on disorder-to-ordered and ordered-to-disordered transitions (Reproduced with permission, Uversky 2015b)

the region (Dunker et al. 2001). This flexibility allows for the reduced time necessary to find the correct docking position and may act as a shock absorber and are more likely to be found in proteins involved in transcription and cell signaling (Dunker et al. 2001; Blackledge 2010).

Table 3.1 A brief list of proteins predicted to contain Long Disordered Regions (LDR) (Reproduced with permission, Romero et al. 1998)

SW ID	Protein description	Seq. length	Max LDR	Disorder location	Avg. strength
SANT_PLAFW	s-Antigen-plasmodium	640	576	51–627	0.99
SANT_PLAF7	s-Antigen-plasmodium	593	490	68–557	0.93
VG48_HSVSA	Hypothetical gene 48 protein from herpes virus saimiri	797	308	413–720	0.96
MLH_TETTH	Micronuclear linker histone poly protein from tetrahymena	633	278	173–450	0.96
CYCLI_HUMAN	Human cyclin (fragment)	598	255	238–492	0.98
NFH_MOUSE	Neurofilament triplet h protein	1087	239	523–761	0.92
RTOA_DICIDI	Slime mold rtoa protein	400	227	77–303	0.96
SR75_HUMAN	Human pre-mRNA splicing factor, srp75	494	225	186–410	0.96
RPBI_CRIGR	Chinese hamster DNA-directed RNA polymerase: largest subunit	1970	221	1610–1830	0.98
LSTP_STAST	Staphylococcus staphyloyticus lysostaphin precursor	480	168	56–223	0.92
YHFI_SALTY	Salmonella typhimurium, hypothetical protein (orf3)	416	154	57–210	0.99
T2FA_DROME	Drosophila melanogaster transcription factor iif, α-subunit	577	153	246–398	0.97
H1_PEA	Garden pea histone H1	265	152	100–252	0.97
110K_PLAKN	Plasmodium knowlesi, 110 kD antigen (fragment)	296	148	135–283	0.91
XYNA_RUMFL	Bifunctional endo-1,4 β-xylanase precursor from ruminoccus flavefaciens	954	129	248–376	0.90
VIT2_CHICK	Vitellogenin ii precursor from chick	1850	125	1142–1266	0.96
SNWA_DICIDI	Snwa protein from slime mold	685	124	398–521	0.97
FHL1_YEAST	Pre-RNA processing protein fhl1 from baker's yeast	936	123	800–923	0.95
HYR1_CANAL	Hyphally regulated protein from Candida albicans (yeast)	937	121	621–741	0.93
ANKB_HUMAN	Ankyrin, brain variant 1 from human	3924	120	1778–1897	0.94

Entropic bristles offer structural support by occupying space and can produce an entropically based repulsion when confined by another protein; binding rates can be also controlled through the motion of the bristle (Fig. 3.12a) (Hoh 1998; Dunker et al. 2001). Entropic bristle proteins like dehydrins have also been used as fusion proteins to help solve solubility issues of recombinant proteins (Fig. 3.12b) by increasing the net charge, decreasing the hydrophobic residue content, and having larger excluded volumes (Santner et al. 2012).

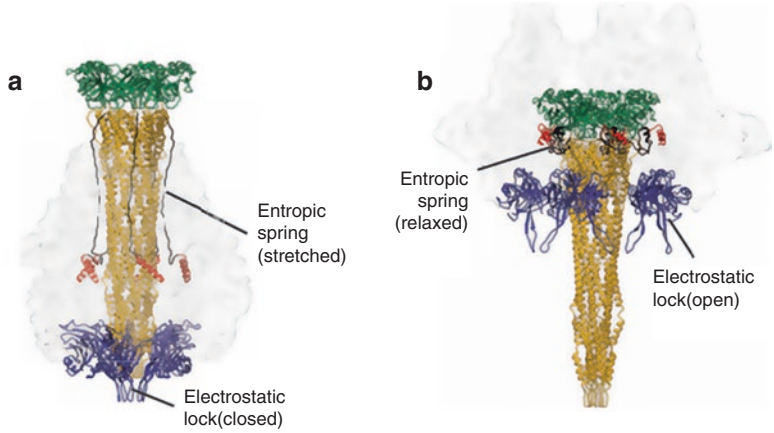

Fig. 3.11 (**a**) The structure of the TcA prepore shows a stretched entropic spring. (**b**) The pore complex of TcA shows a contracted entropic spring (Reproduced with permission, Meusch et al. 2014)

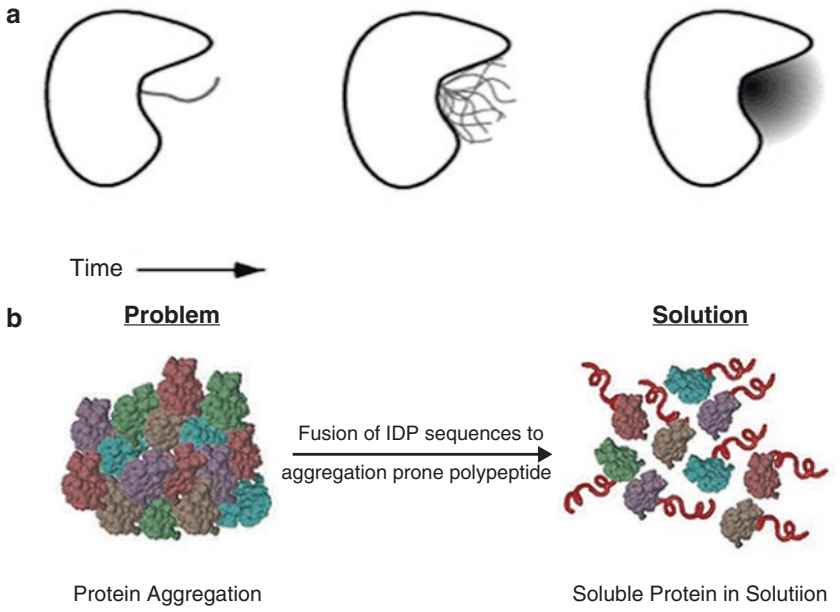

Fig. 3.12 (**a**) Schematic of an entropic bristles as a domain of a protein. Initially the bristle occupies one conformation (arbitrary). Over the timescales of ps to ns, the bristle adopts many different conformations. At the μs timescales, the bristle will occupy all the possible conformations and is represented as a "fuzzy" region to illustrate all conformations (Reproduced with permission, Hoh 1998). (**b**) Schematic representation of how entropic bristles can solve protein aggregation problems (Reproduced with permission, Santner et al. 2012)

3.4.2 Protein Modification

Protein sites subject to posttranslational modifications (PTM) such as phosphorylation or protease cleavage are preferentially located in regions of disorder (Kriwacki et al. 1996; Dunker et al. 2002; Iakoucheva et al. 2004). Phosphorylation has been shown to be overrepresented in disordered regions (Iakoucheva et al. 2004; Gao et al. 2010). In a study by Gao and Xu, they concluded that many of the PTM preferentially occurred in regions of disorder, including: phospho-serine/-threonine/-tyrosine, hydroxylation, sulfotyrosine, S-geranylgeranyl cysteine, deamidated glutamine, 4-carboxyglutamate, 6′-bromotryptophan, and most of methylation (Gao and Xu 2012). IDP functions can be tightly regulated by PTM; failures to do so can cause misfolding, disruption of signaling pathways, or toxic buildup of the proteins (Uversky et al. 2014).

3.4.3 Molecular Assembly

The primary benefit to IDPs, when assembling larger complexes, is the ability for the proteins to overcome steric constraints and gain either local or global structure upon binding (Tompa 2005; Uversky 2010). A more rigid ordered protein would not be able to be assembled with other rigid proteins without some reorganization of domain structure (Namba 2001). As complexes are assembled, disordered proteins can start to form a loosely packed formation which can then range from wrapping around the whole target or forming a more interweaved complex and then tighten to form the final complex as proteins undergo a disordered-to-ordered transition (Namba 2001; Dunker et al. 2002; Uversky 2010). One such example is the binding of SNAP-25, which is initially disordered, to botulinum neurotoxin type A. The SNAP-25, initially in a fully-disordered state, binds to the toxin at distant regions to the active site and wraps around the protein (Washbourne et al. 1997; Breidenbach and Brunger 2004). This example will be discussed in further detail in a later section. Disordered proteins are associated in the assembly of ribosomes, the cytoskeleton, transcription pre-initiation complex, in chromatin, among others (Tompa 2005). Table 3.2 provides a summary of the types of IDP complexes formed with examples for each (Uversky 2010).

3.4.4 Molecular Recognition

Molecular recognition IDPs can be classified as either effectors or scavengers (Sickmeier et al. 2007). Effectors bind to their substrate or ligand using a disorder-to-order transition while scavengers bind small molecules such as metals, also through a disorder-to-order transition (Sickmeier et al. 2007). Cell signaling is the

Table 3.2 Binding modes for IDP complexes

Binding modes for IDP complexes	Description	Examples
Surface-bound molecular recognition features (*MoRFs*)	Short IDR which undergoes a disorder-to-order transition upon binding	α-MoRFs; β-MoRFs; i-MoRFs
Flexible wrappers	An IDP wrapping around its ordered binding partner. The IDP typically have ordered segments linked with disordered regions. These ordered segments bind to remote regions on the ordered binding partner	Calastatin; calpain; inhibitor 2-protein phosphatase 1; $p27^{Kp1}$-cyclin-dependent kinase 2-cyclin A; β-cantenin-T cell factor
Chameleons	Adopt a variety of structures in the bound form and can bind to multiple targets	C-terminus of p53
Penetrators	IDP which penetrates deep into the binding partner upon binding	Ribosomal protein s12 with rRNA
Huggers	Oligomers of IDPs in which monomers gain structure upon oligomerization	Met repressor; Trp repressor; tetramerization domain of p53; tetramerization domain of p73
Coiled-coil-based complexes	Structure of up to seven α-helices intertwined together. Typically contain a repeat pattern of amino acids (*hxxhcxh*, *h* is hydrophobic, *c* is charged and *x* is any amino acid)	
Intertwined strings	Individual α-helices wrapped around each other. Form either left-handed coiled-coils (heptad residue repeat) or right-handed coiled-coils (undecatad residue repeat)	CT670 protein; p21-activated kinase interacting exchange factor; human surfactant protein D
Long cylindrical containers	Multichain coiled-coils which form a hollow cylinder containing a continuous axial pore with the binding capacity for hydrophobic compounds	Cartilage oligomeric matrix protein; GCN4-pAA heptamer
Connectors	Framework molecule for higher order oligomers	ZapA; Bcr-Abl oncoprotein; HSBP1
Armature	Framework on which more complex structures are built	Ectodomain 2
Tweezers and forceps	Transcription factors involved in interaction with DNA	Fos and Jun classes of proteins
Grabbers	Forms a protein scaffold that integrates signaling pathways and regulates post synaptic structures	βPIX coiled-coil-Shank PDZ complex
Tentacles	Structurally consists of a double β-barrel assembly and has six coiled-coils extending from the structure. The hydrophobic patches on the coiled-coils are used for the multi-valent binding of nonnative proteins	Prefoldin
Pullers	Provide a mechanical force to help pull aggregates apart	ClpB
Stackers/β-arcs	Amyloid-like fibrils which are formed by a quaternary structure of cross-β-sheets. β-arcs form a strand-turn-strand motif in which the side chains, not the polypeptide backbone, as in typical β-hairpins	β-amyloid; Sup35p; Ure2P; α-synuclein
Cloud contacts	Form a series of low affinity binding contacts in which the IDP remains disordered. The IDP does not form to a single binding site, rather a "binding cloud"	

Reproduced with permission, Uversky (2010)

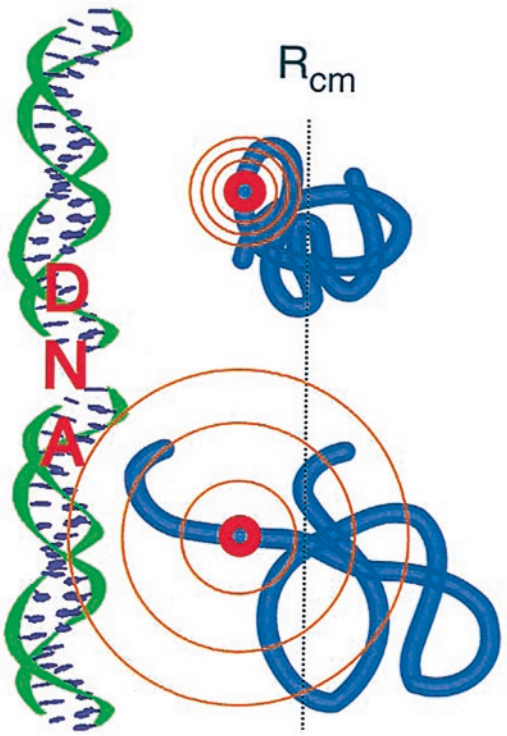

Fig. 3.13 An illustration of the "fly casting" mechanism. When a protein approaches the binding site from a distance of R_{cm}, DNA in this case, the IDP offers a larger capture radius compared to the folded structure. The initial contacts made by the IDP are weak, but they allow the protein to "reel" itself into the binding site to complete the process (Reproduced with permission, Shoemaker et al. 2000)

one of the most common pathways reliant on many IDPs. The flexible nature of the proteins allows for one-to-many signaling as well as many-to-one signaling using a low affinity binding approach, well-suited for cell signaling (Sickmeier et al. 2007; Dunker et al. 2002). The specific binding coupled with low affinity yields a rapid association response (Fig. 3.13) to initiate the signaling/binding and then quickly dissociate when finished (Babu et al. 2011).

3.5 Functional Impact of Intrinsically Disordered Proteins

Some proteins exist in their native state with varying degrees of disorder in their structures. This disorder can either be global (molten globule, pre-molten globule) or local and can play an important role in protein stability, inhibition, signaling, or binding. These proteins can either maintain the disorder upon performing its function while some have been shown to become more ordered while performing its specific function. Both forms of disorder have been shown to offer advantages over a more rigid molecule.

Proteins involving disorder at some point during their lifespan can assist in overcoming steric constraints more effectively than static structures due to the malleable

nature of the disordered protein (Wright and Dyson 1999; Dunker et al. 2001). This flexibility can help play a role in proteins associated with protein-protein binding, protein-DNA binding, protein-RNA binding, cell-cycle control, cell signaling and longevity of the proteins.

Structural flexibility can affect the rate of association and dissociation between two partners, a necessity in regulatory systems such as transcription and translation (Dyson and Wright 2005). The rate of association between disordered proteins with a target would more typically be increased as a more flexible protein would be able to initiate binding at more than one point through a series of low affinity binding sites, unlike static structures (Wright and Dyson 1999; Dunker et al. 2002). A loosened, flexible structure allows for binding with both high-specificity and low-affinity via a "fly-casting mechanism" (Shoemaker et al. 2000; Tompa and Csermely 2004; Sickmeier et al. 2007; Varadi et al. 2015). This "fly-casting mechanism" gives the disordered protein a larger capture radius to make contact over several low-affinity binding sites initially and then bind and fold simultaneous (Fig. 3.14). This may be a significant improvement over a more static globular structure which would have a higher affinity, yet a smaller capture radius (Shoemaker et al. 2000). Dissociation rates could be controlled since the bound disordered protein, which could become ordered upon binding, is bound using low-affinity contact points that can be reversed through a conformational or environmental change.

Flexibility of the protein conformation has been shown to aid in both a one-to-many and many-to-one binding approach between proteins and their targets (Sickmeier et al. 2007; Uversky 2013b; Varadi et al. 2015). Unbound proteins with high degrees of flexibility can more easily adapt to a variety of binding partners to perform their function; unbound ordered proteins with similar structures which can obtain flexibility upon binding can allow for several different proteins targeting the same binding partner (Wright and Dyson 1999; Dunker and Obradovic 2001). Intrinsic plasticity is especially useful for signaling and regulatory proteins which must recognize a range of sometimes unrelated protein conformations, such as misfolded proteins.

On a per residue basis, IDPs have a higher interaction surface area compared to more rigid counterparts (Uversky 2013b). By having a large exposed interface, IDPs allow for a smaller required protein size to perform their specific function compared to using a globular protein. Requiring a smaller protein to perform the necessary functions is an energetic benefit to cells in terms of translation and transcription (Gunasekaran et al. 2003).

In terms of the lifespan, disorder could play an important role in the longevity of certain proteins. A more flexible protein is able to adapt many conformations would help it bind to many cell surface molecules to gain entry or target several intracellular binding partners, while potentially delaying degradation through an immune response or ubiquitin tagging due to the flexibility (Uversky et al. 2014). In the case of cell-signaling, a more rapid turnover of the proteins would be more advantageous to maintain the tightly regulated function and a highly exposed IDP would be an easier target of proteolysis (Wright and Dyson 1999; Uversky 2013b). As mentioned before, the smaller size of the disordered proteins could allow cells more freedom to degrade IDPs involved in signaling with less of an energetic impact when more are needed.

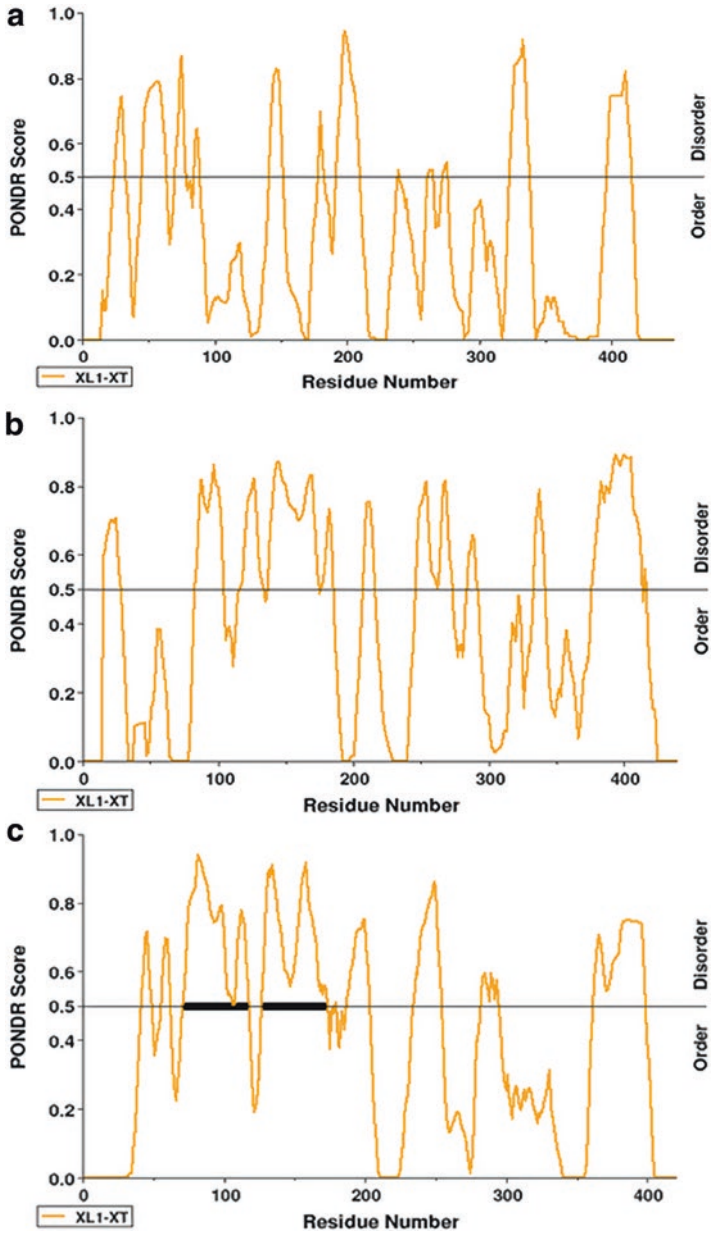

Fig. 3.14 PONDR analysis of botulinum neurotoxin light chains using the XL1-XT predictor for (**a**) type A, 27 % disorder predicted (**b**) type B, 44 % disorder predicted (**c**) type E, 46 % disorder predicted

3.6 Physiological Implications of Intrinsically Disordered Proteins

IDPs have been shown to be largely functionally associated with signaling and recognition pathways, cell-cycle regulation, and gene expression as described earlier. These vital processes are also heavily linked to many types of cancers and disease indicating a role of IDPs (Iakoucheva et al. 2002; Ward et al. 2004; Tompa 2005; Uversky et al. 2014; Cabral et al. 2015; Singh et al. 2015; He et al. 2015; Wu et al. 2015). Disease proteins can be created through genetic mutations that change an amino acid, truncating the protein, or expanding the protein through added amino acids. Mutant proteins can become harmful to the cell if the native function is lost through structural changes or altered binding regions, the ability to be degraded is diminished causing an accumulation of the protein, or a detrimental protein function is created. These mutations can then be potentially inherited by future generations (Uversky et al. 2014).

3.6.1 Intrinsically Disordered Proteins in Neurodegenerative Disorders

Many neurodegenerative diseases can be caused by the misfolding or aggregation of proteins in cells. The involvement of IDPs in these disorders is more common than folded proteins; many of the functional benefits of IDPs can also be the cause of serious damage in cells if the proteins are not able to function properly (Uversky 2015a). The role of IDPs in neurodegenerative disorders is not surprising considering their implication in signaling pathways, regulation, recognition, and the ability to control the activities of their binding partners in which their interactome is highly extended. These processes can be interrupted after mutations to the IDPs or changes to the environment causing a cascade of issues throughout the interactome of that particular IDP (Uversky 2015a). A list of neurodegenerative diseases accompanied by the involved proteins can be seen in Table 3.3 (Uversky 2015a). This table highlights the disordered nature of these proteins along with the extended network of interactions they form.

3.6.2 Intrinsically Disordered Proteins in Cancers

The increased role of IDPs in many cell signaling pathways and as transcription factors makes them an obvious candidate in many types of cancers. For cancers, a study showed 79 % of cancer-associated proteins contained disordered regions of at least 30 residues (Iakoucheva et al. 2002). Mutations to IDPs such as p53 and

Table 3.3 A list of IDPs with their associated disease(s)

Protein (# of residues)	Disease(s)	Disorder by prediction (%)[a]	Number of binding partners on BioGrid[b]
Aβ (42)	Alzheimer's disease Dutch hereditary cerebral hemorrhage with amyloidosis Congophilic angiopathy	16.9 (28.6)	1975 (for the Aβ precursor protein)
Tau (758)	Tauopathies Alzheimer's disease Corticobasal degeneration Pick's disease Progressive supranuclear palsy	77.6 (99.1)	73
Prion protein (231)	Prion diseases Creutzfeld-Jacob disease Gerstmann-Sträussler-Schneiker syndrome Fatal familial insomnia Kuru Bovine spongiform encephalopathy Scrapie Chronic wasting disease	55.8 (61.0)	60
α-Synuclein (140)	Synucleinopathies Parkinson's disease Lewy body variant of Alzheimer's disease	90.7 (37.1)	416
	Diffuse Lewy body disease Dementia with Lewy bodies Multiple system atrophy Neurodegeneration with brain iron accumulation type I		
β-Synuclein (134)	Parkinson's disease Diffuse Lewy body disease	87.3 (52.2)	16
γ-Synuclein (127)	Parkinson's disease Diffuse Lewy body disease	100 (56.8)	26
TDP43 (414)	Amyotrophic lateral sclerosis and frontotemporal lobar degeneration	57.3 (35.8)	286
FUS (526)	Amyotrophic lateral sclerosis	90.7 (72.6)	105
Huntingtin (3144; polyQ tract: 16–37 Qs in norm; >38 Qs in pathology)	Huntington's disease	35.5 (30.4)	193
DRPLA protein (1185; polyQ tract: 7–23 Qs in norm; 49–75 Qs in pathology)	Hereditary dentatorubral-pallidoluysian atrophy	89.5 (84.2)	98

(continued)

Table 3.3 (continued)

Protein (# of residues)	Disease(s)	Disorder by prediction (%)[a]	Number of binding partners on BioGrid[b]
Androgen receptor (919; polyQ tract: 15–31 Qs in norm; 41–81 Qs in pathology)	Kennedy's disease or X-linked spinal and bulbar muscular atrophy	53.9 (46.7)	219
Ataxin-1 (816; polyQ tract: 6–39 Qs in norm; 41–81 Qs in pathology)	Spinocerebellar ataxia 1 Neuronal intranuclear inclusion disease	76.6 (73.4)	254
Ataxin-2 (1312; polyQ tract: 22–31 Qs in norm; >32 Qs in pathology)	Spinocerebellar ataxia 2	93.8 (76.9)	44
Ataxin-3 (376; polyQ tract: 12–40 Qs in norm; 55–84 Qs in pathology)	Spinocerebellar ataxia 3	52.1 (47.1)	76
P/Q-type calcium channel α1A subunit (2505; polyQ tract: 4–16 Qs in norm; 21–28 Qs in pathology)	Spinocerebellar ataxia 6	53.0 (49.3)	94
Ataxin-7 (892; polyQ tract: 4–35 glutamines in norm; 36–306 glutamines in pathology)	Spinocerebellar ataxia 7	89.5 (70.2)	83
TATA-box-binding protein (339; polyQ tract: 25–42 glutamines in norm; >42 glutamines in pathology)	Spinocerebellar ataxia 17	53.9 (52.5)	145
Glial fibrillary acidic protein (1493)	Alexander's disease	82.4 (68.5)	33
DNA excision repair protein ERCC-6 (1493)	Cockayne syndrome	56.8 (47.8)	40
Survival motor neuron protein (294)	Spinal muscular atrophy	69.7 (60.2)	186

Reproduced with permission, Uversky (2015a)

[a]Disorder was predicted by two predictors, PONDR® VSL2 and VLXT (given in parenthesis), respectively. PONDR® VSL2 was chosen because of its high accuracy, whereas PONDR® VLXT was chosen because this predictor was shown to be very sensitive for the presence of molecular recognition features, which are disordered polypeptide segments that are predicted to acquire secondary structure upon forming complexes with binding partners

[b]Interactivity of neurodegeneration-related proteins was evaluated by BioGrid

BRCA1 (breast cancer type 1 susceptibility protein) are associated with most types of cancers due to the large signaling network, especially with transcription factors controlling cell cycle regulation and apoptosis (Uversky et al. 2008; Uversky et al. 2014). In the case of p53, approximately 70 % of the interactions are mediated by the IDR located within the protein while most of the 50+ interactions of BRCA1 occur at the disordered regions (79 % disorder in BRCA1).

3.6.3 Intrinsically Disordered Proteins as Drug Targets

IDPs are an interesting target for therapeutics because of their overrepresentation in cell signaling pathways and disease (cancer, neurodegenerative, cardiovascular, diabetes) (Iakoucheva et al. 2002; Metallo 2010; Babu et al. 2011; Uversky 2015a). Targeting IDPs offers distinct advantages over targeting ordered proteins because of the increased surface and interface areas per residue which acts more dynamically than ordered proteins (Dunker and Uversky 2010; Wang et al. 2011; Zhang et al. 2015). However, as of now, only a few studies have been successful in inhibiting IDPs (Chène 2003; Srinivasan et al. 2004; Erkizan et al. 2009; Hammoudeh et al. 2009; Sievers et al. 2011; Yuzwa et al. 2008; Zhang et al. 2015).

Disrupting the binding of an IDP to its target can be accomplished in several ways with small molecules to decrease the bioavailability: bind to the interface of the IDP binding target (IDP involved drug development), bind to the disordered region of the IDP to prevent binding (IDP targeted drug development), or, in the case of IDPs which gain structure before binding, prevent that refolding and therefore binding (IDP targeted drug development) (Chéne 2003; Babu et al. 2011; Zhang et al. 2015). Decreasing the bioavailability of the IDPs has been shown to be a viable target using regulatory enzymes which post-translationally modify IDPs (Yuzwa et al. 2008; Powers et al. 2009; Puca et al. 2010; Babu et al. 2011).

There is hope that IDPs are a more viable druggable target than ordered proteins and studies have shown some success, as summarized in Table 3.4. This was characterized by showing an average probability of predicted druggable cavities in IDPs was 9 % compared to 5 % in ordered proteins (Zhang et al. 2015). With IDP targeted/involved drug development in the early stages, there are still significant issues. Unlike drug development using ordered proteins, which has well established strategies, IDPs pose challenges due to the lack of a defined 3D structure and dynamic flexibility (Dunker and Uversky 2010; Zhang et al. 2015). Novel approaches to these issues will be needed to take full advantage of the potential IDP drug targets.

3.7 Toxins as Intrinsically Disordered Proteins

3.7.1 Botulinum Neurotoxin

Botulinum neurotoxin (BoNT) is a member of a family of clostridial neurotoxins (CNTs) and is the most toxic substance known with a mouse LD_{50} on the order of 10^{-1} ng/kg body weight (Montecucco and Schiavo 1993; Cai et al. 1999, 2006; Schiavo et al. 2000; Binz et al. 2010). BoNTs are produced by the anaerobic, Gram-positive, spore forming Clostridium bacteria and include seven serotypes labeled as A–G which contain exclusive immunological properties and similar

Table 3.4 Examples of IDP drug targets with their therapeutic importance

Drug target	Therapeutic importance	Inhibitor	Inhibitor target
p53-MDM2	p53 is the most commonly mutated gene in human cancer. Activating the p53 pathway would be an important therapeutic solution	Nutlin	Ordered partner
c-Myc-Max	C-Myc is a deregulated transcription factor associated with human cancers	Peptidomimetic inhibitors	IDP
		10058-F4 and 10074-G5	IDP
EWS-Fli1	EWS-Fli1 is an oncogenic fusion protein involved in Ewing's sarcoma family tumors	YK-4-279	IDP
AF9-AF4	AF9 is a mixed lineage leukemia (MLL) fusion protein causing oncogenic transformation of hematopoietic cells. The AF4 target is the most common fusion protein in acute leukemia	Peptide PFWT	IDP
		Non-peptide compounds	IDP
PTP1B	PTB1B is a negative regulator of insulin and leptin signaling and a therapeutic target of diabetes, obesity, and breast cancer	MSI-1436	Disordered terminus
Aggregating IDPs	Involved in neurodegenerative diseases such as Alzheimer's disease, Down's syndrome, Parkinson's disease, and prion diseases	Molecular tweezers	IDP
		Non-natural amino-acid peptide D-TLKIVW	Amyloid form of IDP
		ELN484228	IDP

Reproduced with permission, Zhang et al. (2015)

pharmacological properties (Harvey et al. 2002; Sharma and Singh 2004; Fujinaga et al. 2004). Each of these serotypes causes botulism; the result of intoxication is flaccid muscle paralysis by blocking the release of acetylcholine at neuromuscular junctions (Singh 2000).

BoNTs consist of a 50 kDa light chain (LC) and a 100 kDa heavy chain (HC) linked through a disulfide bond (Montecucco and Schiavo 1995). Upon binding specifically to presynaptic nerve membrane, BoNT is internalized through endocytosis, and the LC is translocated through a membrane pore formed by the translocation domain (TD) of the HC (Li and Singh 2000). The LC acts as an endopeptidase with remarkable substrate specificity requiring a substantially long peptide sequence of up to 50 amino acids, depending on the serotype (Segelke et al. 2004). This is unique to BoNTs as other microbial metalloproteases can recognize sequences as short as a dipeptide (Silvaggi et al. 2007). Serotype A cleaves SNAP-25 (25-kDa synaptosome associated protein) and the light chain of BoNT/A will be examined in this paper (Li and Singh 1999; Kukreja and Singh 2007; Barash and Arnon 2013).

3.7.1.1 BoNT LCs as Intrinsically Disordered Proteins

Based on disorder predictors such as PONDR (XL1-XT; PONDR score of ≥0.5 is considered disordered) and FoldIndex, the BoNT LC, types A, B, and E, do show regions of disorder (Figs. 3.14 and 3.15) with up to 46% predicted in the case of LCE. The LC fragment of BoNT/A (LCA) and BoNT/B (LCB) both exists in a

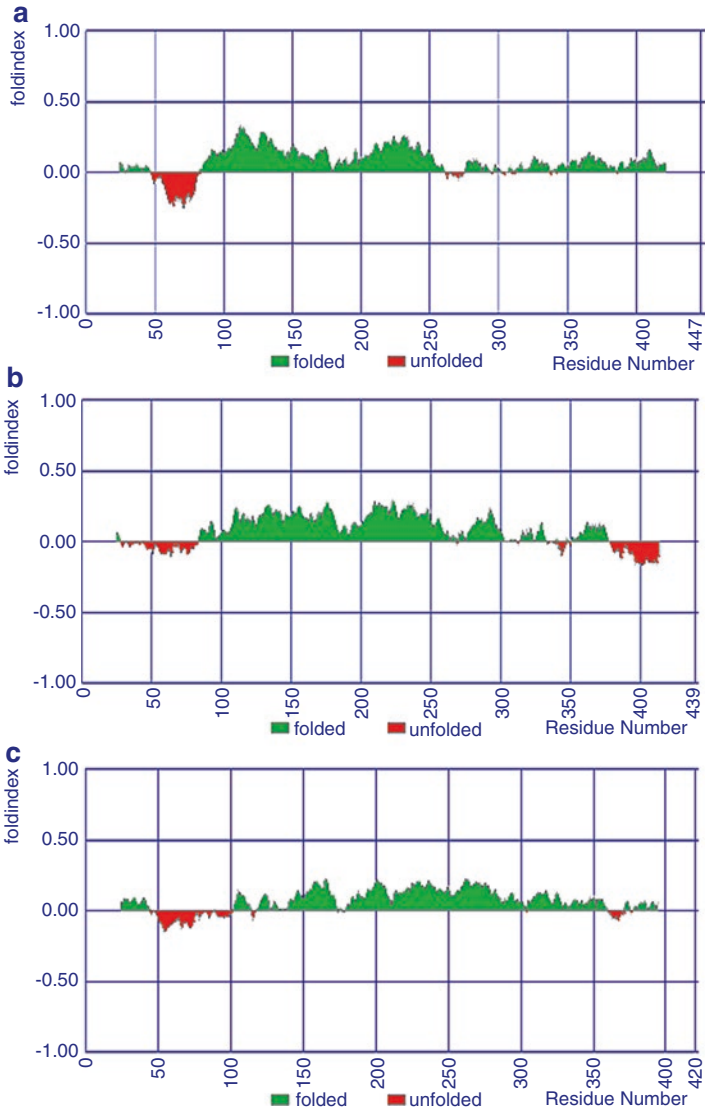

Fig. 3.15 FoldIndex analysis of botulinum neurotoxin light chains for (**a**) BoNT/A LC, (**b**) BoNT/B LC, and (**c**) BoNT/E LC

molten-globule-like state called the pre-imminent molten globule enzyme conformation (PRIME) at 37 °C as well as a molten globule between 50 and 53 °C for LCA and at 47 °C for LCB (Kukreja and Singh 2005; Kumar et al. 2014).

CD spectroscopy initially characterized the LCA, LCB, and LCE over temperature where the T_M for the secondary structure was 48 °C, 52 °C, and 54 °C respectively; the tertiary structure melting temperatures of LCA, LCB, and LCE were 47 °C respectively (Kukreja and Singh 2005; Kumar et al. 2014). This allows for a small window where there is still secondary structure present while the tertiary structure has collapsed for each LC. At 37 °C, both LCA and LCB exhibit an intermediate state where the secondary structure is very similar to the native state and the tertiary structure is still largely present, however, is less compact allowing for greater conformational flexibility.

ANS binding of LCA, LCB, and LCE were observed over temperature with the highest ANS binding found at 50 °C, 47 °C, and 47 °C, respectively, and indicates the proteins existing in a molten globule state (Kumar et al. 2014). In the case of LCA and LCB, there is a small deviation in the ANS binding trend observed at 37 °C which could represent a unique conformation formed in which the tertiary structure is loosened with the secondary structure remaining in-tact. Interestingly, this correlates with both the biologically relevant temperature of the LC and the temperature of optimum activity (Kukreja and Singh 2005; Kumar et al. 2014).

The use of UV/Vis spectroscopy and fluorescence also indicated the presence of a molten globule structure and PRIME structure in the case of LCA and LCB. The a/b ratios of LCA and LCB both showed an intermediate at 37 °C, indicative of increased Tyr exposure (Kukreja and Singh 2005; Kumar et al. 2014). This increase in Tyr exposure is evidence of a refolding of the tertiary structure as the temperature is raised to the biologically relevant temperature. The fluorescence emission spectrum of LCA becomes red-shifted as temperature is increased; at 20 °C the emission is 318 nm and shifts to 329 nm at 60 °C. This shift in emission is characteristic of the Trp becoming more exposed to a polar environment.

The LCA was also characterized under different pH conditions which showed the formation of a molten globule conformation (Cai et al. 2006). The secondary structure of LCA remained unchanged when pH was 4.5, 5.0, 5.5, and 7.0. The Trp lifetime was longer as pH decreased which suggests a loosely packed tertiary structure with a high amount of secondary structure. ANS binding was higher at pH 4.7 compared to pH 7.0 (Li and Singh 2000) while there was a sharp increase observed between pH 5.5 to pH 5.0 (Cai et al. 2006). These observations indicated a molten globule is formed at low pH. This conformation change was also observed using anisotropy; anisotropy indicates an elongated shape at lower pH which may play a significant role the translocation of LCA across the membrane.

Along with a globally loosened structure at the physiological temperature, the BoNT/A LC has a highly flexible C-terminus loop spanning the final 24 residues which has prevented researchers from crystalizing the full-length LC (Mizanur et al. 2013). In place of the full-length LC, crystal structures have been solved for truncated versions of the LC including a 424-residue LC (PDB ID: 2IMC; Silvaggi et al. 2007), a 424-residue LC with a small SNAP-25 peptide (PDB ID: 3DDA;

Kumaran et al. 2008), and a 424-residue LC bound with the sn2 domain of SNAP-25 (PDB ID: 1XTG; Breidenbach and Brunger 2004) among many others. Since BoNT proteins recognize the tertiary structure of their respective substrate and due to this not only binding is affected with respect to size of substrate or enzyme, endopeptidase activity is also severely affected (Baldwin et al. 2004).

According to theoretical predictions, BoNT LCs (types A, B, and E) have significant amount of disorder present in their structure (Figs. 3.14 and 3.15). The solution structures of LCA and LCB at 25 °C appear to be ordered, but the optimally active conformations at 37 °C appear to be less ordered or partially disorder, known as the PRIME conformation (Kukreja and Singh 2005; Kumar et al. 2014). LCE did not clearly show the presence of a PRIME conformation, however, it exists in a molten globule conformation at 47 °C and is optimally active at that temperature with its full-length substrate (Kumar et al. 2014). Above observations indicated the presence of intrinsically disordered conformation in BoNT LCs, which is in accordance with their requirement of long substrate.

3.7.1.2 BoNT/A as a Molten Globule

In its optimally active form, the BoNT/A holotoxin and the BoNT/A endopeptidase have been shown to exist as molten globule structures (Cai et al. 1999; Kukreja and Singh 2005). The critical factor for the molten globule structure for the holotoxin is the reduction of the disulfide bond between the heavy chain (HC) and the light chain (LC), while pH and temperature are responsible for the molten globule structure of the LC.

In a study done by Cai and Singh (2001), BoNT/A was shown to exist in a molten globule structure when the disulfide bond between the HC and LC was reduced, however, when this bond was intact, the toxin remained in a folded form. The non-reduced form of BoNT/A showed no structural differences based on far- and near-UV CD between 20 and 37 °C and had a T_M of 53 °C for both the secondary and tertiary structure which showed a cooperative unfolding. This is a distinct difference from the reduced BoNT/A as the reduced toxin loses almost all tertiary structure by 37 °C (T_M of 35 °C) with only small amount of secondary structure lost (T_M of 44 °C), each with a non-cooperative unfolding pattern. This 9 °C difference in the T_M of the secondary and tertiary structure allows for a significantly loosened structure with a high degree of flexibility.

The state of the hydrophobic structure was monitored using ANS binding. An ANS binding ratio for both the non-reduced and reduced toxin was first calculated in which the reduced form could bind 110 molecules of ANS while the non-reduced sample could only bind 50 molecules. The non-reduced toxin showed peak ANS binding at 53 °C where there is almost no secondary or tertiary structure which is typical for a non-molten globule protein. The reduced toxin showed highest levels of ANS binding at 35–37 °C, a temperature which coincided with a lack of tertiary structure but the presence of a high amount of the secondary structure, typical of a protein with a molten globule type conformation.

This molten globule structure found for the reduced toxin around 35–37 °C also showed maximum activity with the full-length SNAP-25 substrate (tagged with GST). This molten globule structure may facilitate in the binding of SNAP-25 as this process is very selective. Assuming the BoNT molecule remains in a molten globule state upon substrate binding, a maximally active molten globule is unique to BoNT and other proteins have yet been discovered with this feature; however, one alternative explanation is BoNT undergoing a disorder-to-order transition upon binding of SNAP-25 and returning to a molten globule state once in its unbound state. In either case, the BoNT molecule relies on disorder for its activity.

3.7.1.3 Longevity of BoNT Intoxication

One of the most unusual characteristics of BoNT is the longevity of action on cells which is one of the features that has made it such a versatile therapeutic. Each serotype shows a varying lifetime of action with BoNT/A having the longest duration in humans (typically 3–6 months but up to 1 year for some conditions) (Scott 2004; Blumenfeld et al. 2004; Glogau 2004). While BoNT/B and /E show considerably shorter lifetimes than BoNT/A they remain very potent and can survive for weeks to months with type E showing the shortest lifetime among the three types mentioned (Foran et al. 2003). Since each serotype has been shown to have a flexible structure (PRIME structure for LCA and LCB; optimally active molten globule for LCE) and the enzymes having a prolonged lifetime of action, the question of whether their flexible nature is responsible for this longevity arises.

As discussed earlier, a more flexible protein would increase its survival in cells by delaying degradation (Uversky et al. 2014). Some of the proposed mechanisms for BoNT survival include Tyr-phosphorylation (Ferrer-Montiel et al. 1996), increased resistance to ubiquitination (Raiborg and Stenmark 2009), palmitoylation of cysteine residues (Montal 2010), and S-nitrosylation (Montal 2010). Each of these pathways may be affected by the varying flexibility of the LCs and the longevity differences between serotypes may be due to those differences.

3.7.2 Other Toxins as IDPs and the Importance of Disorder in Their Biological Function

The DisProt database is currently made up of nearly 700 proteins; at least 14 of those are toxins, or toxin associated proteins (Table 3.5). Other protein toxins, including BoNT/A, BoNT/B, BoNT/E, tetanus toxin, diphtheria toxin, and anthrax lethal factor, were analyzed using PONDR (XL1-XT) and showed significant amounts of disorder (Fig. 3.16). With such wide ranging advantages of disorder in the proteome, toxins would also benefit from IDRs. Many toxins function by

Table 3.5 Summary of example toxins/toxin associated proteins identified in the DisProt database

Toxin protein (DisProt ID)	Source organism	Protein size (amino acids)/ disorder (%)	Disorder region 1 role	Disorder region 1 functional class/functional subclass	Disorder region 2 role	Disorder region 2 functional class/functional subclass	Disorder region 3 role	Disorder region 3 functional class/functional subclass	Toxin description
Salt-mediated killer toxin 1 beta chain (DP00180_C003)	*Picia farninosa* (*yeast*)	77 (100 %)	Fully disordered protein gains function through disorder-to-order transition	Molecular assembly/ protein-protein binding					Kills sensitive strains of yeast
Cholera enterotoxin subunit A (DP00250)	*Vibrio cholerae*	258 (4 %)	11 residue region (Activation loop; 44–54) near the C-terminal gains function through order-to-disorder transition. Modified by limited proteolysis and/or reduction of disulfide bond	Entropic chain/ autoregulatory	10 residue region (Active site loop; 66–75) undergoes an order-to-disorder transition. Flexible loop allows substrates to enter active site to undergo ADP-ribosylation	Unknown/ autoregulatory	11 residue region (KDEL binding region; 245–255) undergoes an unknown transition. Elongated structure extends through the B5 pore of cholera toxin subunit B to bind to KDEL receptor	Unknown/ protein-protein binding	Catalyzes ADP-ribosylation of Gs alpha to activate adenylate cyclase. Leads to overproduction of cAMP and hypersecretion of chloride, bicarbonate, and water

(continued)

Table 3.5 (continued)

Toxin protein (DisProt ID)	Source organism	Protein size (amino acids)/ disorder (%)	Disorder region 1 role	Disorder region 1 functional class/functional subclass	Disorder region 2 role	Disorder region 2 functional class/functional subclass	Disorder region 3 role	Disorder region 3 functional class/functional subclass	Toxin description
Perfringolysin (DP00280)	*Clostridium perfringens*	500 (10%)	28 residue segment (TMH1; 190–217) undergoes order-to-order transition proceeds through disordered intermediate	Unknown/ unknown	24 residue segment (TMH2; 288–311) undergoes order-to-order transition proceeds through disordered intermediate	Unknown			
Colicin-E9 (DP00342)	*Escherichia coli*	582 (14%)	83 residue segment (1–83) on N-terminus undergoes disorder-to-order transition as part of translocation of toxin	Molecular assembly/ protein-protein binding					Targets *E. coli* and related bacteria. Enters cell through translocation using disordered N-terminus. Once in the cell, the toxin acts as an endonuclease on both dsDNA and ssDNA

Leukotoxin (DP00345)	*Pasteurella haemolytica*	953 (7 %)	70 residue segment (884–953) on C-terminus undergoes disorder-to-order transition	Molecular recognition effectors/ protein-protein binding					Exotoxin that attack host leukocytes, causing cell rupture. Once bound to host LFA-1 integrin, a signaling cascade is induced causing Tyr phosphorylation of CD18 tail, elevated intracellular Ca^{2+} levels, and lysis of host cell
Diptheria toxin repressor (DP00374)	*Corynebacterium diphtheriae* (strain PW8)	226 (8 %)	17 residue segment (130–146) remains in disordered state when functional	Entropic chain/ flexible linker, spacer					Iron-binding repressor of Dtx gene expression. Under iron excess, represses ripA
Hemolysin, chromosomal (DP00389)	*Escherichia coli*	1023 (6 %)	61 residue segment (963–1023) on C-terminus undergoes disorder-to-order transition	Unknown/ protein-protein binding					Exotoxin that attack blood cell membrane, cause cell rupture via unknown mechanisms

(continued)

Table 3.5 (continued)

Toxin protein (DisProt ID)	Source organism	Protein size (amino acids)/ disorder (%)	Disorder region 1 role	Disorder region 1 functional class/functional subclass	Disorder region 2 role	Disorder region 2 functional class/functional subclass	Disorder region 3 role	Disorder region 3 functional class/functional subclass	Toxin description
Calmodulin-sensitive adenylate cyclase (DP00395)	*Bacillus anthracis*	800 (2 %)	14 residue segment (578–591) undergoes disorder-to-order transition	Modification site/substrate, ligand binding					One of three proteins that compose anthrax toxin. Infects many mammalian species. Calmodulin-dependent adenylyl cyclase that causes edema when associated with PA. Causes increase in intracellular cAMP levels in host

Ras-related C3 botulinum substrate 1 (Isoform A) (DP00408)	*Homo sapiens*	192 (23 %)	9 residue region (Switch I, 30–38) undergoes disorder-to-order transition	Molecular assembly/ protein-protein binding	17 residue region (Switch II, 59–75) undergoes disorder-to-order transition	Molecular assembly/ protein-protein binding	19 residue region (Insertion, 76–94) undergoes an unknown transition	Unknown/ unknown	Small GTPase associated with PM cycles between active GTP-bound and inactive GDP-bound states. Active state binds to variety of effector proteins to regulate several cellular responses
Colicin-N (DP00461)	*Escherichia coli*	387 (8 %)	30 residue segment (Translocation domain; 40–69) undergoes disorder-to-order transition. Imports protein across outer membrane. Protein then binds to third domain of TolA protein, possibly OmpF protein too	Molecular recognition effectors and molecular assembly/ protein-protein binding					Channel forming colicin that depolarize the cytoplasmic membrane, cellular energy is dissipated. Produced and active against *E. coli* and similar bacteria

(continued)

Table 3.5 (continued)

Toxin protein (DisProt ID)	Source organism	Protein size (amino acids)/ disorder (%)	Disorder region 1 role	Disorder region 1 functional class/functional subclass	Disorder region 2 role	Disorder region 2 functional class/functional subclass	Disorder region 3 role	Disorder region 3 functional class/functional subclass	Toxin description
Bifunctional hemolysin/ adenylate cyclase (DP00591)	*Bordatella pertussis*	1706 (41%)	701 residue region (Repeat in Toxin (RTX); 1006–1706) undergoes disorder-to-order transition upon binding to calcium	Unknown/ substrate, ligand binding					Causes whooping cough by elevating cAMP-concentrations to disrupt normal cell function. IDR is unfolded in absence of calcium; binding to calcium induces secondary and tertiary structure formation
Thermostable direct hemolysin 2 (DP00668)	*Vibrio parahaemolyticus*	189 (6%)	11 residue region (1–11) on N-terminus undergoes unknown transition upon becoming functional	Unknown/ unknown					Exotoxin that attack blood cell membrane, cause cell rupture via unknown mechanisms

Type II secretion protein M (DP00725)	*Vibrio cholerae* serotype O1	165 (28 %)	42 residue region (N-terminal of eriplasmic domain; 44–85) remains disordered in functional state	Unknown/ unknown	5 residue loop (148–152) undergoes unknown transition	Unknown/ unknown			Involved in protein export pathway. Required by cholera toxin for secretion across outer membrane
Botulinum neurotoxin type E light chain (DP00732)	*Clostridium botulinum*	421 (5 %)	11 residue loop (234–244) with unknown transition	Unknown/ unknown	10 residue region (C-terminal; 412–421) remains disordered in functional state	Unknown/ unknown			BoNT inhibits neurotransmitter release in neuromuscular junctions by cleaving SNAP-25

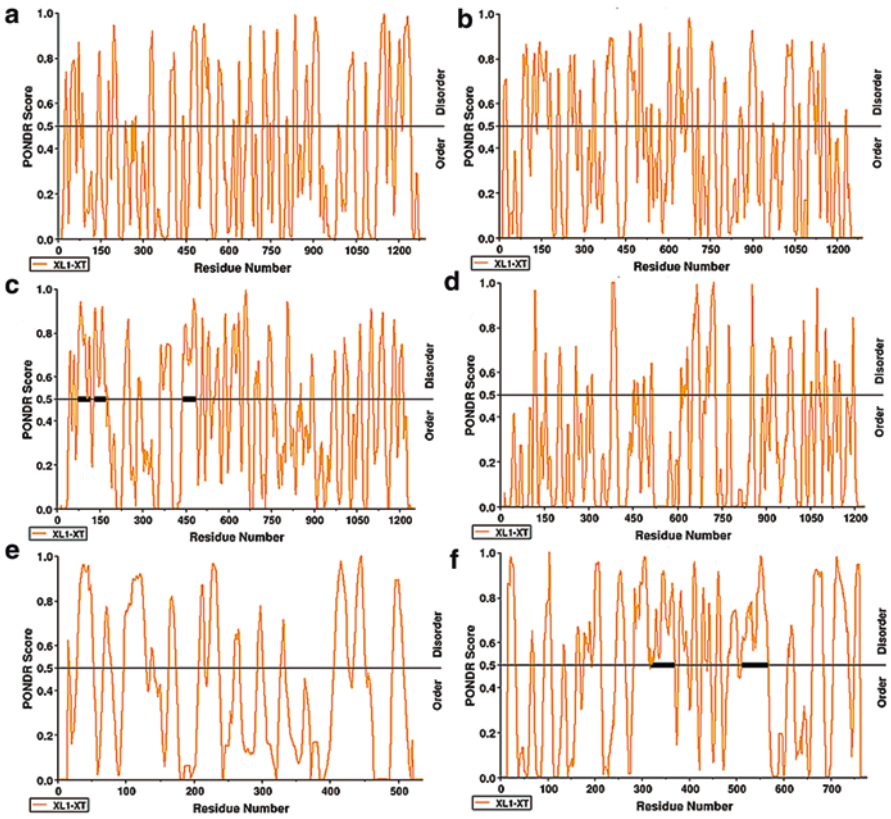

Fig. 3.16 PONDR prediction (XL1-XT) for different protein toxins and their percent disorder. Shown are (**a**) BoNT/A (34.59 %), (**b**) BoNT/B (37.60 %), (**c**) BoNT/E (40.65 %), (**d**) Tetanus toxin (21.43 %), (**e**) Diphtheria toxin (37.20 %), and (**f**) Anthrax lethal factor (53.35 %)

disrupting cell processes with the end results often times being cell or organism death. Obviously, cells would try to prevent damage from foreign toxins which would then force toxins to find ways around the cell protection.

A more flexible toxin able to adapt many conformations increases its chances to bind to cell surface molecules to gain entry past the first line of defense for cells, the plasma membrane (PM) (Uversky et al. 2014). Once bound to a receptor on the PM or endosome, many toxins, both ordered and disordered, have mechanisms of translocating through the PM; some of these toxins use disordered regions to assist in their translocation, such as colicin-E9 and colicin-N.

Once inside the cells, disorder gives proteins the freedom to adopt many different conformations which could give toxins an advantage in selectively recognizing substrates. Flexible toxins could more readily identify and optimize binding to cellular targets using this disorder. Long-lasting toxins, such as BoNT, could be explained by

the flexible nature of the proteins. Toxins could delay degradation through an immune response or change in ubiquitin tagging due to the flexible nature (Uversky et al. 2014).

3.8 Conclusions

Intrinsic disorder in proteins is a relatively recently discovered field of protein structure research that has reshaped the way researchers view the structure-function relationship of proteins. The number of articles on this topic has continued to grow since the first papers were published in the early-1980s, and the topic was largely unrecognized until the mid-1990s (Sickmeier et al. 2007; Wright and Dyson 2015). Prediction tools developed have enhanced the discovery of disorder within proteins by providing convenient methods to study a large number of proteins in shorter periods of time. By first predicting the disorder, a more targeted approach to examining the disorder with experiments can be developed.

Continued technical advances in both experimental techniques and predictors have allowed for further progress in understanding the mechanisms behind IDPs. The limitations of techniques such as X-ray crystallography with IDPs hinder the characterization of proteins with larger amounts of disorder. NMR and fluorescence based techniques offer great hope in furthering the understanding of this class of proteins and improvements to these techniques have allowed for increased sensitivity, larger proteins to be studied, and characterization in natural cellular systems (Berlow et al. 2015; Wright and Dyson 2015). A further understanding of the mechanisms behind IDPs will offer great insight into protein folding, protein binding/interactions, cell signaling, and other cellular processes with IDPs now clearly playing a vital role in the proteome.

As more studies are performed it has become clear that IDPs are heavily involved in important cellular processes such as signaling and recognition pathways, cell-cycle regulation, and gene expression. The flexible nature of these proteins allow for regulation of association/dissociation constants, life cycle of the protein, and provide a mechanism for both many-to-one and one-to-many binding. Due to the cellular dependence on these proteins, many diseases are also involved with IDPs making IDPs an interesting target for therapeutics (Berlow et al. 2015; Wright and Dyson 2015).

The relationships between complexity and amount of IDPs in the organism coupled with structural flexibility associated with IDPs qualify this class of proteins a better candidate for evolutionary studies. IDPs in the primordial world may have offered an evolutionary advantage to organisms because of their versatility (Vamvaca et al. 2004); a more versatile protein would favor organisms by decreasing the necessary genomic size and energy costs needed to produce the proteins needed to survive. These proteins can be used as a model for evolutionary theory, which fulfils the requirement of adaptability and natural selection. Careful study of this class of proteins can provide us a better evolutionary theory and possibly be used as therapeutic starting points where these flexible proteins can be altered to better solve current catalytic needs.

References

Antz C, Geyer M, Fakler B et al (1997) NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. Nature 385(6613):272–275

Babu MM, Lee RV, Groot NS et al (2011) Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol 21(3):432–440

Baldwin MR, Bradshaw M, Johnson EA, Barbieri JT (2004) The C-terminus of botulinum neurotoxin type A light chain contributes to solubility, catalysis, and stability. Protein Expr Purif 37:187–195

Barash JR, Arnon SS (2013) A novel strain of Clostridium botulinum that produces type B and type H botulinum toxins. J Infect Dis 209(2):183–191

Berlow RB, Dyson HJ, Wright PE (2015) Functional advantages of dynamic protein disorder. FEBS Lett 589:2433–2440

Binz T, Sikorra S, Mahrhold S (2010) Clostridial neurotoxins: mechanism of SNARE cleavage and outlook on potential substrate specificity reengineering. Toxins (Basel) 2(4):665–682

Blackledge M (2010) Mapping the conformational mobility of multidomain proteins. Biophys J 98(10):2043–2044

Blumenfeld AM, Dodick DW, Silberstein SD (2004) Botulinum neurotoxin for the treatment of migraine and other primary headache disorders. Dermatol Clin 22(2):167–175

Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. Chem Rev 106(8):3055–3079

Bracken C (2001) NMR spin relaxation methods for characterization of disorder and folding in proteins. J Mol Graph Model 19(1):3–12

Braselmann E, Chaney JL, Clark PL (2013) Folding the proteome. Trends Biochem Sci 38(7):337–344

Breidenbach MA, Brunger AT (2004) Substrate recognition strategy for botulinum neurotoxin serotype A. Nature 432(7019):925–929

Cabral KM, Raymundo DP, Silva VS et al (2015) Biophysical studies on BEX3, the p75NTR-associated cell death executor, reveal a high-order oligomer with partially folded regions. PLoS One 10(9):e0137916

Cai S, Singh BR (2001) Role of the disulfide cleavage induced molten globule state of type A botulinum neurotoxin in its endopeptidase activity. Biochemistry 40(50):15327–15333

Cai S, Sarkar HK, Singh BR (1999) Enhancement of the endopeptidase activity of botulinum neurotoxin by its associated proteins and dithiothreitol. Biochemistry 38(21):6903–6910

Cai S, Kukreja R, Shoesmith S et al (2006) Botulinum neurotoxin light chain refolds at endosomal pH for its translocation. Protein J 25(7–8):455–462

Chen JW, Romero P, Uversky VN et al (2006) Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. J Proteome Res 5(4):888–898

Chène P (2003) Inhibiting the p53–MDM2 interaction: an important target for cancer therapy. Nat Rev Cancer 3(2):102–109

Dolgikh D, Gilmanshin R, Brazhnikov E et al (1981) α-lactalbumin: compact state with fluctuating tertiary structure? FEBS Lett 136(2):311–315

Dolgikh D, Kolomiets A, Bolotina I et al (1984) 'Molten-globule' state accumulates in carbonic anhydrase folding. FEBS Lett 165(1):88–92

Dolgikh D, Abaturov L, Bolotina I et al (1985) Compact state of a protein molecule with pronounced small-scale mobility: bovine-lactalbumin. Eur Biophys J 13(2):109–121

Dunker AK, Obradovic Z (2001) The protein trinity—linking function and disorder. Nat Biotechnol 19(9):805–806

Dunker AK, Uversky VN (2010) Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. Curr Opin Pharmacol 10(6):782–788

Dunker A, Lawson J, Brown CJ et al (2001) Intrinsically disordered protein. J Mol Graph Model 19(1):26–59

Dunker AK, Brown CJ, Lawson JD et al (2002) Intrinsic disorder and protein function. Biochemistry 41(21):6573–6582
Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6(3):197–208
Eliezer D, Chiba K, Tsuruta H et al (1993) Evidence of an associative intermediate on the myoglobin refolding pathway. Biophys J 65:912–917
Erkizan HV, Kong YL, Merchant M et al (2009) A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. Nat Med 15(7):750–756
Evans JNS (1995) Biomolecular NMR spectroscopy. Oxford University Press, Oxford
Ferrer-Montiel AV, Canaves JM, Dasgupta BR et al (1996) Tyrosine phosphorylation modulates the activity of clostridial neurotoxins. J Biol Chem 271(31):18322–18325
Fiaux J, Bertelsen EB, Horwich AL et al (2002) NMR analysis of a 900K GroEL–GroES complex. Nature 418(6894):207–211
Fontana A, Polverino de Laureto P, De Filippis V (1993) Molecular aspects of proteolysis of globular proteins. In: van den Tweel W, Harder A, Buitelear M (eds) Protein stability and stabilization. Elsevier Science, Amsterdam
Fontana A, Zambonin M, Polverino de Laureto P et al (1997) Probing the conformational state of apomyoglobin by limited proteolysis. J Mol Biol 266:223–230
Fontana A, Polverino de Laureto P, Spolaore B et al (2004) Probing protein structure by limited proteolysis. Acta Biochim Pol 51:299–321
Foran PG, Mohammed N, Lisk GO et al (2003) Evaluation of the therapeutic usefulness of botulinum neurotoxin B, C1, E, and F compared with the long lasting type A. J Biol Chem 278(2):1363–1371
Fujinaga Y, Inoue K, Watarai S et al (2004) Molecular characterization of binding subcomponents of Clostridium botulinum type C progenitor toxin for intestinal epithelial cells and erythrocytes. Microbiology 150:1529–1538
Gao J, Xu D (2012) Correlation between posttranslational modification and intrinsic disorder in protein. Pac Symp Biocomput 17:94–103.
Gao J, Thelen JJ, Dunker AK, Xu D (2010) Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. Mol Cell Proteomics 9(12):2586–2600
Glogau RG (2004) Treatment of hyperhidrosis with botulinum toxin. Dermatol Clin 22:177–185
Gunasekaran K, Tsai C-J, Kumar S et al (2003) Extended disordered proteins: targeting function with less scaffold. Trends Biochem Sci 28(2):81–85
Hammoudeh DI, Follis AV, Prochownik EV et al (2009) Multiple independent binding sites for small molecule inhibitors on the oncoprotein c-Myc. J Am Chem Soc 131:7390–7401
Harvey SM, Sturgeon J, Dassey DE (2002) Botulism due to Clostridium baratii type F toxin. J Clin Microbiol 40(6):2260–2262
He Y, Chen Y, Mooney SM et al (2015) Phosphorylation-induced conformational ensemble switching in an intrinsically disordered cancer/testis antigen. J Biol Chem 290(41):25090–25102
Hemmings HC Jr, Nairn AC, Aswad DW et al (1984) DARPP-32, A dopamine- and adenosine 3′:5′-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. J Neurosci 4(1):99–110
Hernàndez MA, Avila J, Andreu JM (1986) Physiochemical characterization of the heat-stable microtubule-associated protein MAP2. Eur J Biochem 154:41–48
Hoh JH (1998) Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. Proteins 32:223–228
Hubbard SJ, Eisenmenger F, Thornton JM (1994) Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. Protein Sci 3:757–768
Iakoucheva LM, Kimzey AL, Masselon CD et al (2001) Identification of intrinsic order and disorder in the DNA repair protein XPA. Protein Sci 10:560–571
Iakoucheva LM, Brown CJ, Lawson JD et al (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323:573–584

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 32(3):1037–1049

Ishima R, Torchia DA (2000) Protein dynamics from NMR. Nat Struct Biol 7(9):740–743

Kikhney AG, Svergun DI (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett 589:2570–2577

Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. J Magn Reson 241(100):74–85

Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. PNAS 93:11504–11509

Kukreja RV, Singh BR (2005) Biologically active novel conformation state of botulinum, the most poisonous poison. J Biol Chem 280(47):39346–39352

Kukreja RV, Singh BR (2007) Comparative role of neurotoxin-associated proteins in the structural stability and endopeptidase activity of botulinum neurotoxin complex types A and E. Biochemistry 46:14316–14324

Kumar R, Kukreja RV, Cai S et al (2014) Differential role of molten globule and protein folding in distinguishing unique features of botulinum neurotoxin. Biochim Biophys Acta 1844:1145–1152

Kumaran D, Rawat R, Ahmed SA et al (2008) Substrate binding mode and its implication on drug design for botulinum neurotoxin A. PLoS Pathog 4:e1000165

Li L, Singh BR (1999) Structure-function relationship of clostridial neurotoxins. J Toxicol Toxin Rev 18(1):95–112

Li L, Singh BR (2000) Spectroscopic analysis of pH-induced changes in the molecular features of type A botulinum neurotoxin light chain. Biochemistry 39:6466–6474

Li X, Romero P, Rani M et al (1999) Predicting protein disorder for N-, C- and internal regions. Genome Inform Ser Workshop Genome Inform 10:30–40

Linding R, Jensen LJ, Diella F et al (2003) Protein disorder prediction: implication for structural proteomics. Structure 11:1453–1459

Marsh JA, Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. Biophys J 98:2383–2390

Metallo SJ (2010) Intrinsically disordered proteins are potential drug targets. Curr Opin Chem Biol 14:481–488

Meusch D, Gatsogiannis C, Efremov RG et al (2014) Mechanism of Tc toxin action revealed in molecular detail. Nature 508:61–65

Mizanur RM, Frasca V, Swaminathan S et al (2013) The C terminus of the catalytic domain of type A botulinum neurotoxin may facilitate product release from the active site. J Biol Chem 288(33):24223–24233

Montal M (2010) Botulinum neurotoxin: a marvel of protein design. Annu Rev Biochem 79:591–617

Montecucco C, Schiavo G (1993) Tetanus and botulinum neurotoxins: a new group of zinc metalloproteases. Trends Biochem Sci 18:324–327

Montecucco C, Schiavo G (1995) Structure and function of tetanus and botulinum neurotoxins. Q Rev Biophys 28:423–472

Namba K (2001) Roles of partly unfolded conformations in macromolecular self-assembly. Genes Cells 6:1–12

Obradovic Z, Peng K, Vucetic S et al (2003) Predicting intrinsic disorder from amino acid sequence. Proteins 53:566–572

Ohgushi M, Wada A (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains. FEBS Lett 164:21–24

Pancsa R, Tompa P (2012) Structural disorder in eukaryotes. PLoS One 7(4):e34687

Parthasarathy S, Murthy MRN (2000) Protein thermal stability: insights from atomic displacement parameters (*B* values). Protein Eng 13(1):9–13

Powers ET, Morimoto RI, Dillin A et al (2009) Biological and chemical approaches to diseases of proteostasis deficiency. Annu Rev Biochem 78:959–991
Prilusky J, Felder CE, Zeev-Ben-Mordehai T et al (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21(16):3435–3438
Ptitsyn OB (1987) Protein folding: hypotheses and experiments. J Protein Chem 6:273–293
Puca R, Nardinocchi L, Givol D et al (2010) Regulation of p53 activity by HIPK2: molecular mechanisms and therapeutical implications in human cancer cells. Oncogene 29:4378–4387
Raiborg C, Stenmark H (2009) The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. Nature 458:445–452
Romero P, Obradovic Z, Kissinger CR et al (1997) Identifying disordered regions in proteins from amino acid sequences. Proc IEEE Int Conf Neural Networks 1:90–95
Romero P, Obradovic Z, Kissinger CR et al (1998) Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput 3:437–448
Romero P, Obradovic Z, Li X et al (2001) Sequence complexity of disordered protein. Proteins 42(1):38–48
Santner AA, Croy CH, Vasanwala FH et al (2012) Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. Biochemistry 51:7250–7262
Schiavo G, Matteoli M, Montecucco C (2000) Neurotoxins affecting neuroexocytosis. Physiol Rev 80:717–766
Scott AB (2004) Development of botulinum toxin therapy. Dermatol Clin 22:131–133
Segelke B, Knapp M, Kadkhodayan S et al (2004) Crystal structure of Clostridium botulinum neurotoxin protease in a product-bound state: evidence for noncanonical zinc protease activity. PNAS 101(18):6888–6893
Sharma SK, Singh BR (2004) Enhancement of the endopeptidase activity of purified botulinum neurotoxins A and E by an isolated component of the native neurotoxin associated proteins. Biochemistry 43:4791–4798
Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. PNAS 97(16):8868–8873
Sickmeier M, Hamilton JA, LeGall T et al (2007) DisProt: the database of disordered proteins. Nucleic Acids Res 35:D786–D793
Sievers SA, Karanicolas J, Chang HW et al (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nature 475:96–100
Silvaggi NR, Boldt GE, Hixon MS et al (2007) Structures of Clostridium botulinum neurotoxin serotype A light chain complexed with small molecule inhibitors highlight active-site flexibility. Chem Biol 14:533–542
Singh BR (2000) Intimate details of the most poisonous poison. Nature 7(8):617–619
Singh S, Colonna G, Di Bernardo G et al (2015) The gene expression profiling of hepatocellular carcinoma by a network analysis approach shows a dominance of intrinsically disordered proteins (IDPs) between hub nodes. Mol Biosyst 11:2933–2945
Srinivasan RS, Nesbit JB, Marrero L et al (2004) The synthetic peptide PFWT disrupts AF4-AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells. Leukemia 18:1364–1372
Tompa P (2002) Intrinsically unstructured proteins. Trends in Biochem Sci 10:527–533
Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett 579:3346–3354
Tompa P (2011) Unstructural biology coming of age. Curr Opin Struct Bio 21:419–425
Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 37(12):509–516
Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. FASEB J 18:1169–1175
Trombitás K, Greaser M, Labeit S et al (1998) Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. J Cell Biol 140:853–859
Urbauer JL, Cowley AB, Broussard HP et al (2015) Solution structure and properties of AlgH from Pseudomonas aeruginosa. Proteins 83:1137–1150

Uversky VN (2002a) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11:739–756
Uversky VN (2002b) What does it mean to be natively unfolded? Eur J Biochem 269:2–12
Uversky VN (2010) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40:1623–1634
Uversky VN (2013a) Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta 1834:932–951
Uversky VN (2013b) A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci 22:693–724
Uversky VN (2015a) Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. Front Aging Neurosci 7:1–6
Uversky VN (2015b) The multifaceted roles of intrinsic disorder in protein complexes. FEBS Lett 589:2498–2506
Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37:215–246
Uversky VN, Gillespie JR, Fink AL (2000) Why are "Natively Unfolded" proteins unstructured under physiological conditions? Proteins 41:415–427
Uversky VN, Davè V, Iakoucheva LM (2014) Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. Chem Rev 114:6844–6879
Varadi M, Zsolyomi F, Guharoy M, Tompa P (2015) Functional advantages of conserved intrinsic disorder in RNA-binding proteins. PLOS ONE 10(10):e0139731
Vamvaca K, Vögeli B, Kast P et al (2004) An enzymatic molten globule: efficient coupling of folding and catalysis. PNAS 101(35):12860–12864
Vucetic S, Obradovic Z, Vacic V et al (2005) DisProt: a database of protein disorder. Bioinformatics 21(1):137–140
Wang J, Cao Z, Zhao L et al (2011) Novel strategies for drug discovery based on intrinsically disordered proteins (IDPs). Int J Mol Sci 12:3205–3219
Ward JJ, Sodhi JS, McGuffin LJ et al (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337:635–645
Washbourne P, Pellizzari R, Baldini G et al (1997) Botulinum neurotoxin types A and E require the SNARE motif in SNAP-25 for proteolysis. FEBS Lett 418:1–5
Weinreb PH, Zhen W, Poon AW et al (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 35(43):13709–13715
Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331
Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signaling and regulation. Nat Rev Mol Cell Biol 16(1):18–29
Wu HY, Chen SF, Hsieh JY et al (2015) Structural basis of antizyme-mediated regulation of polyamine homeostasis. PNAS 112(36):11229–11234
Xue B, Li L, Meroueh SO et al (2009) Analysis of structured and intrinsically disordered regions of transmembrane proteins. Mol BioSyst 5(12):1688
Yuzwa SA, Macauley S, Heinonen JE et al (2008) A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau *in vivo*. Nat Chem Biol 4(8):483–490
Zandany N, Lewin L, Nirenberg V et al (2015) Entropic clocks in the service of electrical signaling: 'Ball and chain' mechanisms for ion channel inactivation and clustering. FEBS Lett 589:2441–2447
Zhang Y, Cao H, Liu Z (2015) Binding cavities and druggability of intrinsically disordered proteins. Protein Sci 24:688–705

Chapter 4
Implication of Molten Globule on the Function and Toxicity of a Protein

Harkiran Preet Kaur Dhaliwal, Raj Kumar, and Bal Ram Singh

Keywords Molten globule (MG) • Native state • Cellular processes • Ligand binding • Enzyme activity • Chaperones • Translocation • Toxins • Botulinum neurotoxin • Diphtheria toxin • Anthrax toxin • Genetic diseases

4.1 Introduction

Molten globule (MG) states are characterized by the presence of significant amount of native secondary structure and compactness in overall protein structure (Dolgikh et al. 1981; Ohgushi and Wada 1983, 1984; Bychkova et al. 1988). Molten globules (MGs) also known as partially denatured states, lack rigid tertiary structure and possess loosely packed hydrophobic core with higher accessibility of hydrophobic surface area to solvent, as demonstrated in earlier reports utilizing different biophysical techniques such as microcalorimetry, circular dichroism (CD), fluorescence (polarization and lifetime measurements), nuclear magnetic resonance (NMR), deuterium exchange mass spectroscopy, and others (Dolgikh et al. 1985; Gil'manshin and Ptitsyn 1987; Semisotnov et al. 1987) for the study of MG structure. MGs are either defined as equilibrium state or as kinetic intermediates (Gil'manshin and Ptitsyn 1987; Bychkova et al. 1988), observed during the unfolding transition of protein structure. The equilibrium MG state is either observed at an acidic pH (~3.6), an alkaline pH or at an intermediate concentration of a strong denaturant such as ~2 M guanidinium chloride, where the MG state is stably populated. The examples include bovine carbonic anhydrase (BCA) (Wong and Tanford 1973; Henkens et al. 1982; Dolgikh et al. 1983; Ptitsyn et al. 1983), alpha-lactalbumin (Dolgikh et al. 1981; Gil'manshin et al. 1982), bovine pancreatic ribonuclease A (Denton et al. 1982), cytochrome c (Ohgushi and Wada 1983; Naeem and Khan 2004), and apomyoglobin (Eliezer et al. 1997). In the case of globular proteins (such as cytochrome c and apomyoglobulin), where alkaline or acidic environment induced denaturation leads to the substantial formation of unfolded state without the formation of

R. Kumar, B.R. Singh, *Protein Toxins in Modeling Biochemistry*, SpringerBriefs in Biochemistry and Molecular Biology, DOI 10.1007/978-3-319-43540-4_4

intermediate, the salt-induced refolding from the unfolded state is known to cause the formation of equilibrium intermediate MG state. Alcohol, sodium dodecyl sulfate (SDS), hydrostatic pressure, and weak salt denaturants are some of the other agents which are known to cause the formation of the MG state (Iram and Naeem 2012; Marion et al. 2015; Rahaman et al. 2015). Proteins which do not demonstrate equilibrium MG state are commonly known to unfold as a kinetic intermediate (Dolgikh et al. 1984; Kuwajima et al. 1985; Gil'manshin and Ptitsyn 1987; Semisotnov et al. 1987; Craig et al. 1987; Chen et al. 2008; Nishimura et al. 2008). Examples include the states of proteins which are trapped by chaperones right after their biosynthesis or before they become fully folded (Ptitsyn et al. 1995; Bychkova and Ptitsyn 1993; Bychkova et al. 1988) and the intermediates that are produced as a result of point mutations eventually preventing polypeptides from complete folding (Ptitsyn et al. 1995; Bychkova and Ptitsyn 1995).

The ability of the MG state to retain the memory of its native structure makes it suitable for existence in the constantly changing living cell environment. The progress made in recent years for a better understanding of MG states has enabled scientists to establish the role of these structures in biological processes. The establishment of the cellular role was first suggested by the group of Ptitsyn in 1988 (Bychkova et al. 1988). The study became a paradigm for the establishment of functional roles for MG states at the cellular level. Examples of biological functions of the MG state includes, the recognition of proteins by chaperones (Martin et al. 1991; van der Vies et al. 1992), to facilitate release or interaction of protein ligands (Bychkova et al. 1992; Uversky and Narizhneva 1998; Uversky 2003), involvement in the interaction or penetration of proteins into cell membranes or organelles (van der Goot et al. 1991, 1992; Bychkova et al. 1996), the transfer of retinal from its bloodstream carrier to the cell-surface receptor (Bychkova et al. 1992, 1998), and the pore-forming domains of toxins or proteins that act as carriers of large hydrophobic ligands which form MG states to fulfill their functions (Bychkova and Ptitsyn 1993; Ptitsyn 1995; Uversky and Dunker 2010). The present chapter enlists details about the roles of the MG state in the biological processes. The physiological roles of MG structures can be divided into five major groups—(1) functional native molten globule, (2) role in cellular processes, (3) enzymatic activity, (4) translocation of toxins/proteins across the membranes, and (5) genetic diseases (Fig. 4.1). The understanding of the functions of MGs in biological processes is of utmost importance in order to study the relevance of structure–function relationship of MGs in the physiological system. The study of MGs act as a bridge between the biophysical (structural) and biological sciences in order to understand many unexplained mechanisms of cellular processes as well as of diseases. Overall, the present chapter will help in the understanding of two important aspects of protein chemistry—(1) the structural and mechanistic role of the MG state in protein folding, and (2) the impact of the MG conformation in the mechanism of action of proteins/toxins in the cellular environment.

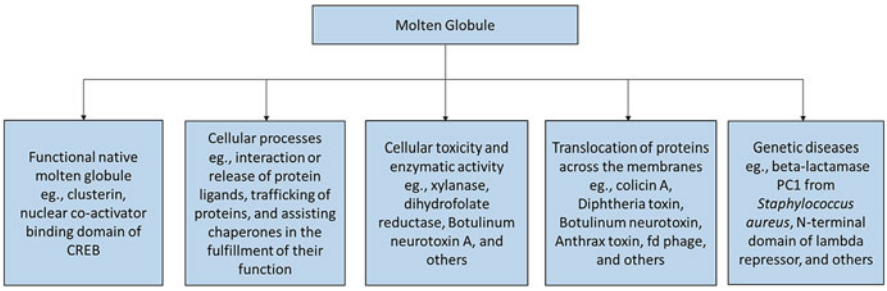

Fig. 4.1 Schematics of MG and its functions

4.2 Functional Native MG

There are certain proteins which are known to possess domains in their native structure that exist in the MG state. The examples of such proteins include clusterin, the nuclear co-activator binding domain (NCBD) of cyclic adenosine monophosphate (cAMP)-responsive element-binding protein (CREB), toxoplasma P2, a ribosomal protein produced by *Toxoplasma gondii*, and nucleosome core particle.

4.2.1 Clusterin

Clusterin is a multifunctional glycoprotein that constitutes a native molten globular region and an ordered domain. Clusterin, also known as sulfated glycoprotein-2, was first found as a component of secretion from Sertoli cells (Sertoli 1865). These cells are found in close association with developing germ cells within the testis. Sertoli cells are known to nurse the germ cells by providing nutrients and regulatory factors (Griswold 1995). Clusterin function is similar to small heat shock chaperone proteins, that binds to misfolded or stressed proteins to prevent stress-induced aggregation or keep them in a certain folded state which is suitable for binding to other proteins (Poon et al. 2000; Trougakos and Gonos 2006). Clusterin exists in both intracellular and extracellular chaperone forms, with the latter one being better understood (Trougakos and Gonos 2006).

The structural characterization of clusterin has been unsuccessful owing to its property of aggregation and sticking to hydrophobic surfaces. However, such properties also indicate towards the existence of molten globular domain in the native structure of clusterin. The currently known structure of clusterin is based on computational modeling. The mature secreted clusterin (rat) protein consists of two subunits, α (34 kDa) and β (47 kDa), linked via five disulfide bonds (Bailey et al. 2001) (Fig. 4.2). There are six N-linked glycosylation sites unevenly distributed

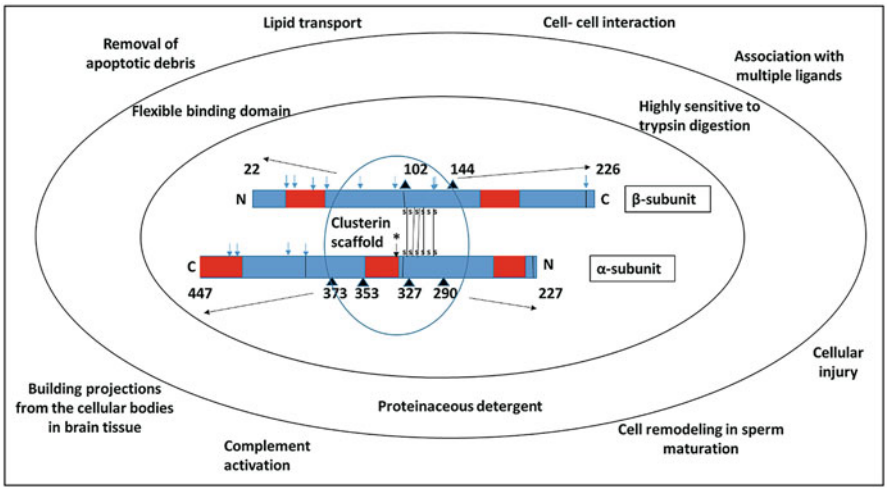

Fig. 4.2 Schematic representation of processes leading to the formation of clusterin MG structure and its functions. The *black triangles* denote N-glycosylation, **s---s**—disulfide bond, *blue arrows*—trypsin cleavage site, *red boxes*—putative amphipathic helix and *blue boxes*—predicted disordered regions. '*' denotes resistant trypsin digestion site within the predicted disordered region in ANS-bound clusterin. The clusterin (rat) structure is based on the predictions and experiments conducted by Bailey et al. (2001). Clusterin scaffold in the center is indicated by a *blue circle* including disulfide bond sites and N-glycosylation sites. The four arms are represented by four *black arrows* directing outward from the scaffold including disordered and ordered structures. The disordered clusterin structure with amphiphilic regions forms the MG structure which helps in binding to various ligands to exert its functions (Reproduced with permission, Bailey et al. 2001)

among the two subunits. Bailey et al. (2001) performed sequence analysis experiment for disordered structure prediction that showed the presence of a MG like structure in the three long disordered regions. The N-terminus of β-subunit shows the presence of longest disordered region and the C-terminus of α-subunit shows the next longest region (Fig. 4.2). The disordered and ordered regions exist in opposite order to each other in the two subunits, providing polarity of disorder and order structure. The disordered clusterin regions were found to be highly conserved among rat, bovine, and human species. The amphipathic alpha-helices present in the long, flexible regions of both the subunits provide multiple dynamic binding sites for amphipathic ligands. The disordered regions were also demonstrated to contain ~80 % of the active trypsin digestion sites, indicating towards the possibility of the existence of a MG like conformation in the predicted disordered structure. Clusterin is known to bind to various ligands, such as proteins, lipids, and peptides. Its property of binding to 8-anilinonaphthalene-1-sulfonic acid (ANS), a commonly used hydrophobic probe was effective in conducting binding studies. The effect of urea denaturation on the clusterin-ANS binding was studied by titrating clusterin-ANS complex against increasing concentration of urea (Bailey et al. 2001). The urea titration of ANS fluorescence showed unfolding at a lower urea concentration along

with the absence of typical sigmoidal curve or cooperativity, which are the typical characteristics of MG protein (Yang et al. 2001). The results further strengthened the hypothesis of the existence of MG structure in the predicted disordered regions, when compared to other MG proteins (apomyoglobin, and alpha-lactalbumin). In addition, the ANS-bound clusterin also showed protection of one trypsin digestion site within the predicted disordered region, indicated by '*' in the Fig. 4.2. The findings of this study strongly suggest the presence of a MG state in the ANS binding region of clusterin. As shown in the Fig. 4.2, most of the N-terminal glycosylation sites are present within or close to the disulfide bonds, indicating the formation of a clusterin scaffold in the middle, along with negatively charged carbohydrates localized within the scaffold and the ordered as well as disordered structures with amphiphilic helices extending outward from the scaffold as arms. The amphiphilic helices present on the arms interact with each form to form an antiparallel four-helix bundle.

The clusterin requires binding to a variety of hydrophobic ligands (of different sizes and shapes) for the exertion of its functions such as clearing of cell debris, which could be possible due to the presence of various amphipathic alpha-helices in its structure (Tsuruta et al. 1990; de Silva et al. 1990). However, the structure should also be flexible or at least possess few dynamic sites required for binding to various ligands. Bailey et al. (2001) have also demonstrated the function of clusterin as a proteinaceous detergent (shown by its ability to dissolve bacteriorhodopsin) that has been proposed to help in binding to or maintaining solubility of various hydrophobic complexes and denatured proteins required for the tissue remodeling process (Bailey and Griswold 1999). Overall, the above-mentioned studies suggests that the natively disordered regions along with amphiphilic helices in clusterin could form a flexible MG binding state that could aid in its binding to a variety of ligands, which could help in execution of its diverse cellular functions such as cell remodeling in sperm maturation, building projections from the cell bodies in brain, lipid transport, removal of apoptotic debris, and protective effect against cellular stresses (Bailey and Griswold 1999; Dunker et al. 2001; Bailey et al. 2001).

Another vital function of clusterin is to act as a chaperone, which has been demonstrated by: (1) inhibition of the stress-induced (heat or reduction) irreversible precipitation of different proteins or substrates (for example, undiluted human serum) under *in vitro* as well as *in-vivo* conditions (Humphreys et al. 1999; Poon et al. 2000; Rosenberg et al. 2002) and (2) binding to stressed proteins for formation of solubilized and stable high molecular weight complexes involving an adenosine triphosphate (ATP)-independent mechanism (Poon et al. 2000). This process helps in creating a reservoir of inactive but stable protein complexes which can be later utilized by ATP-dependent chaperones to retrieve functional proteins (Poon et al. 2000). Generally, an elevated temperature is known to increase chaperone activity of the intracellular small heat shock proteins (iHSPs) proteins. For example, mammalian αA-crystallin and yeast HSP26 (form non-aggregated dimeric forms from the oligomeric structure at elevated temperature) (Bova et al. 1997; Haslbeck et al. 1999). However, the increased temperature showed either no effect on disso-

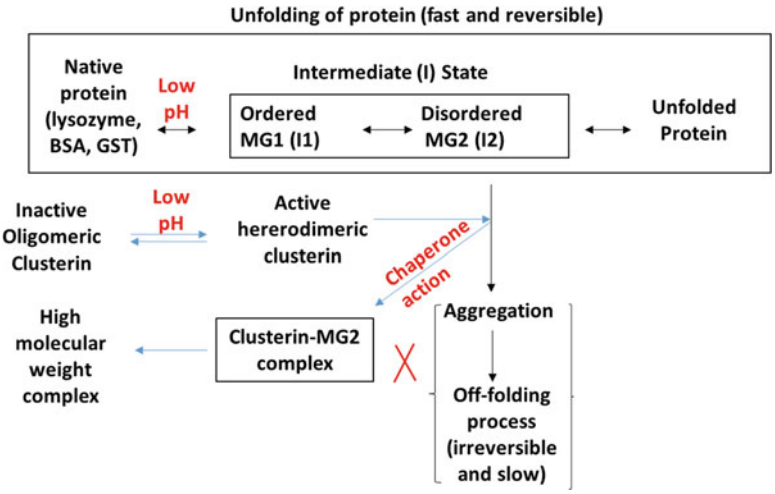

Fig. 4.3 Schematic representation of the chaperone action of clusterin. The native proteins such as lysozyme, BSA or GST at low pH conditions undergo unfolding transitions by forming intermediate ordered (MG1 or I1) and disordered (MG2 or I2) states. The active heterodimeric clusterin produced as a result of low pH exposure forms a complex with MG2 or I2 state of native protein to inhibit its aggregation and further stabilizes the protein by forming high molecular weight complex (Reproduced with permission, Poon et al. 2002a, b)

ciation of clusterin heterogeneous aggregates or very little effect on chaperone activity of clusterin (Poon et al. 2002b). Poon et al. (2002b) showed that the chaperone action of clusterin increases at mildly acidic pH by using CD and bis-ANS binding experiments. Interestingly, clusterin is the only known chaperone to increase its activity under acidic pH conditions. According to the mechanism proposed by Poon et al. (2002a) the acidic pH leads to the unfolding of the native protein (stressed proteins such as, lysozyme, bovine serum albumin (BSA), and glutathione-*S*-transferase (GST)) and hence adopting series of intermediates (MG1/I1 and MG2/I2) as MG state in the proteins (Fig. 4.3). Once the intermediate state exposes enough hydrophobic regions it forms aggregates and enters the irreversible aggregation form by following the slow off-folding pathway. The role of the chaperone-active heterodimeric form of clusterin starts along this pathway where it binds to intermediate form (MG2) of the protein and gets sequestered into large clusterin-intermediate complex (Fig. 4.3). The clusterin aggregates were demonstrated to dissociate at low pH forming more heterodimeric clusterin from the oligomeric clusterin resulting in an increase of its chaperone activity. The authors also proposed that the dissociation at low pH occurs due to the increase of solvent exposed hydrophobicity, as very little secondary or tertiary structure changes were observed in the structure of clusterin at acidic pH. Clusterin presents a unique example of a MG state function, where it not only involves a native MG domain but also in its ability to bind to stressed native MG state proteins in order to inhibit their precipitation.

4.2.2 NCBD

The NCBD of CREB has also been demonstrated to possess a native MG domain. The functions of NCBD include interaction of CREB with number of other proteins such as the steroid receptor co-activators (Li and Chen 1998, p. 53; Gu et al. 1997, p. 73; Burge et al. 2009), the interferon regulatory proteins (IRF) 3 and 7 (Lin et al. 1998, 2000) and the viral protein Tax (Scoggin et al. 2001). NCBD binds majorly to two ligands, (1) antibody-coupled T-cell receptor (ACTR) or (2) interferon regulatory transcription factor (IRF)-3 eventually leading to formation of two varied ligand specific folding states, NCBD,ACTR, and NCBD,IRF3 (Demarest et al. 2002; Waters et al. 2006; Qin et al. 2005). In the ligand-free state, the NCBD forms a compact structure with a high degree of helicity along with the lack of sigmoidicity in the unfolding curve (Demarest et al. 2002, 2004; Ebert et al. 2008) suggesting the existence of native MG state. Kjaergaard et al. (2010) demonstrated that the MG state of NCBD, ACTR complex is similar to the NCBD in a ligand-free state using NMR spectroscopy. However, NCBD, IRF3 was shown to possess a different topology than the ligand-free state. The results indicate that the ACTR is able to selectively bind to a prefolded structure of NCBD existing in an MG state (Kjaergaard et al. 2010). The authors propose that NCBD is able to exist in a number of unfolded or partially folded (MG) states, which in turn aid in binding to structurally diverse ligands.

4.2.3 Toxoplasma P2

Toxoplasma P2, a ribosomal protein, is produced in *Toxoplasma gondii*, a parasite which causes toxoplasmosis. Toxoplasma P2 is a part of the supramolecular assembly in the ribosome and is also known to be involved in extra-ribosomal functions (Mishra et al. 2015). Mishra et al. (2015) have demonstrated for the existence of intrinsically MG state in toxoplasma P2, using gel filtration, biophysical spectroscopic techniques (CD, fluorescence spectroscopy and differential scanning calorimetry (DSC)), and NMR at different pH and temperature conditions. The protein structure showed the presence of a monomer at low pH (2.0) and multimer at higher pH. Interestingly, the structure of Toxoplasma P2 shows the presence of helices and unstructured regions (Mishra et al. 2015) indicating towards the presence of flexible and dynamic MG state. The protein sequence constitutes flexible C-terminal (with MG state) at the one end and helices at the N-terminal end. To summarize, the inherent or native dynamic MG structure of Toxoplasma P2 has been proposed to be responsible for the extra-ribosomal activities of the protein.

4.2.4 Histone Octamer

A native MG state exists in the histone octamer which is a part of nucleosome core particle. The histone octamer is known to be wrapped by 1¾ turns of double-stranded DNA consisting of 146 base pairs. The histone octamer structure is made

up of positively charged tails which are found to be sensitive to protease digestion (Ausio et al. 1989). Remarkably, the charged tails remain undetected in the nucleosome X-ray structure (Luger et al. 1997; Luger and Richmond 1998). The properties of the tail structure suggest the presence of disordered structure in the native state. Hyperacetylation of these nucleosome tails have also been suggested to have a role in DNA replication, (Waterborg and Matthews 1983; Chahal et al. 1980) RNA transcription, (Chahal et al. 1980; Mizzen and Allis 1998), and chromosomal remodeling (Dunker et al. 2001). Electron microscopy scans suggest deformation of the hyperacetylated nucleosomes on the carbon surface as compared to the regular nucleosomes (Oliva et al. 1990). One of the reasons pointed out by the authors for the observance of such difference is that hyperacetylation causes nucleosomes to become much more sensitive to the forces encountered during electron microscopy sample preparation, leading to the deformation of nucleosomes (Oliva et al. 1990). However, hyperacetylation also increases the overall negative charge on the nucleosomes, which leads to overall increase in charge repulsion and reduced rigidity in the nucleosome structure (Dolgikh et al. 1981; Ptitsyn and Uversky 1994; Kuwajim 1996). Dunker et al. (2001) have also reported less rigidity in the structure of hyperacetylated nucleosomes as compared to normal nucleosomes by using (contact mode method) atomic force microscopy (AFM) imaging. AFM imaging has an advantage over the electron microscopy because of its ability to control the amount of force applied on the nucleosomes particles as well as the experiment can be carried out in a hydrated state similar to physiological environment (Dunker et al. 2001). To summarize the above reports, the hyperacetylation in nucleosome core particles has been proposed to cause an increase in negative charge in the nucleosome particles which increases the charge imbalance, similar to as induced by high or low pH eventually leading to a transition to MG state (Dunker et al. 2001).

Not much is known about the mechanisms by which the native MGs bind to their ligands to form fully folded complexes, required for the functioning of proteins in cellular processes. The direct role of native MG in cellular function is yet to be established, but this state is definitely required for proper protein-protein interactions as indicated by the above-mentioned studies, needed for several cellular functions. The presence of native MGs is further supported by the existence of their functioning in the cells, which is a fairly sufficient reason for exploring more about the mechanism of such states.

4.3 Role of MGs in Cellular Processes

There are certain proteins or domains in the protein structure which exist in the native state but they must denature to perform their biological functions. A precise mechanism of how these proteins can exist or denature in the physiological environment is still unclear. However, it could be reasoned out that the presence of nucleoprotein complexes, membrane proteins, cytoskeletal proteins, and organelles in the cellular environment could assist in the existence of MG states (Ptitsyn et al. 1995). Protein denaturation at the membrane surface that aids in MG existence can be explained mainly via two factors (Ptitsyn et al. 1995): (1) The presence of negative

charge near the cell membrane which could possibly attract protons and can lead to a decrease of local pH (van der Goot et al. 1991), (2) The presence of organic moieties near the membranes which could lead to a decrease in the local dielectric constant (Bychkova and Ptitsyn 1993). Hence, the most appropriate experiment to study MG state in the physiological environment involves studying denaturation of a protein in the water-organic mixture at either moderate or low pH.

The MG states of proteins are involved in various cellular processes mainly involving interaction or release of protein ligands, translocation of proteins to specific organelles by remaining in their unfolded states in the cytoplasm or near the membrane surfaces (Randall and Hardy 1989; Hartl and Neupert 1990), and aiding chaperones in the implementation of their function. The examples of the above-mentioned processes are explained below in detail and have been summarized in a tabulated form in Table 4.1.

Table 4.1 Summary of the functions of MG state in cellular processes

Proteins	Molten globule state	Function in molten globule state	References
Retinol-binding protein (RBP)	– At low pH between 5.5. and 3.0 – Moderately low pH and the moderately low dielectric constant	Release of protein ligand (retinol) at low pH conditions	Bychkova et al. (1992, 1998), Bychkova and Ptitsyn (1993)
α-fetoprotein	At pH 3.0 in water solution	Allows release of unsaturated fatty acids and estrogens to embryonal and few cancer cells	Ptitsyn et al. (1995), Uversky et al. (1995)
Alpha-hemolysin (HlyA)	Acylated form of toxins	Requires binding to fatty acids for exertion of hemolytic and cytotoxic effects	Herlax and Bakas (2007)
Plastocyanin	Intermediate protein state during refolding	Electron transfer at the luminal side of the thylakoid membrane of chloroplasts	Li et al. (1990), Koide et al. (1993)
Cytochrome c	Moderately low pH and moderately low dielectric constant	Intermediate in protein folding	Ohgushi and Wada (1983), Bychkova et al. (1996)
Steroidogenic acute regulatory protein (StAR)	Active form of protein present proximal to the membrane	Switch control for cholesterol entry into the mitochondria	Bose et al. (1999)
Nucleosome core particle	Hyperacetylation	DNA replication and RNA transcription	Kuwajim (1996), Dunker et al. (2001)
Ribosomal proteins (for example, HIV-1 Rev, BIV Tat, L11-C76, YP2b)	Possess disordered RNA recognition sites and exists in the absence of RNA	Helps in binding to RNA	Yonath and Franceschi (1997), Surendran et al. (2004), Shojania and O'Neil (2006), Markus et al. (1997), Zurdo et al. (1997)

4.3.1 Ligand Binding

Temperature changes, pH changes, or interaction with ligands or biomolecules are the important factors known to enhance the chances of transition of a native protein structure into an MG state. The release of non-polar ligands upon denaturation of proteins is a crucial feature of MG state. The examples of ligand carriers include *Retinol-binding protein (RBP)*, α-fetoprotein (AFP), Alpha-hemolysin or alpha-toxin (HlyA), and ribosomal proteins.

4.3.1.1 Retinol-Binding Protein (RBP)

The very first report of a protein involved in such mechanism was that of RBP which is a carrier of large non-polar ligand i.e. retinol (Bychkova et al. 1992). Regulation of RBP is crucial mainly because of its function of helping in proper retinol exposure to the embryo during pregnancy and its successful transport at the maternal-fetal interface. The deficit of either one of the proteins is related to early embryo mortality or developmental malformations. Bychkova et al. (1992) reported the release of retinol by RBP at low pH, (5.5–3.0) concurrent with transitioning of protein into the MG state. Ligands, such as retinol, are tightly packed in the core of a rigid protein via non-polar groups, therefore, the existence of MG or denatured state becomes critical where ligands are less tightly packed which could aid in the release of ligand (Bychkova and Ptitsyn 1993). The observance of MG state in RBP has also been shown at moderate pH in water–methanol mixtures (Bychkova et al. 1998). The reports by Bychkova et al. (1998) demonstrated removal of retinol from RBP and denaturation of RBP occurrence by the concerted action of moderately low pH as well as moderately low dielectric constant. The solutions used for the experiments contained a mixture of low ionic strength buffer and methanol in variable proportions at 37 °C (with a range of values expected to exist in the *in vivo* microenvironment). These reports indicated towards the idea of MG existence in the cellular environment and their function in the biological processes.

4.3.1.2 α-fetoprotein (AFP)

Binding of AFP to its non-polar ligand also involves MG or denatured state (Ptitsyn et al. 1995). AFP is also known as alpha-1-fetoprotein, alpha-fetoglobulin, or alpha fetal protein. It is a major plasma oncofetal protein which is produced by the yolk sac and the liver during fetal development (Mizejewski 2004). However, not much is known about the exact biological functions of AFP during fetal and perinatal development. Instead, higher levels of AFP in the serum often serve as a fetal defect marker for detection of cancer predominantly predating in the detection of birth defects, malformations, and congenital anomalies (Tomasi 1977; Mizejewski 2004). AFP is also known to be involved in the transfer of unsaturated fatty acids and

estrogens to embryonal and fewer cancer cells (Abelev 1993). Reports by Uversky et al. (1995) demonstrated the existence of AFP from human cord serum in an MG state at pH 3.1 in water solution with characteristics of MG such as native-like far-ultraviolet (UV) CD spectrum, tryptophan (Trp) fluorescence spectrum, strong ANS binding, and non-cooperative melting temperature. Uversky et al. (1997) demonstrated the release of ligands from AFP while simultaneously causing a significant rearrangement of the protein structure with predominant characteristics of MG. The reports also suggested the existence of ligand-free form in a compact and native-like secondary structure with the absence of rigid tertiary structure. Interestingly, the authors also reported that the native AFP structure was not able to reconstitute by adding back of ligands after the release of ligands or formation of MG state. Overall the above-mentioned reports indicated the possible existence of MG state for the α-fetoprotein, that allows the release of unsaturated fatty acids upon denaturation or transformation into the MG state.

4.3.1.3 Alpha-Hemolysin (HlyA)

HlyA (110 kDa), a pore-forming toxin secreted by pathogenic strains of *Escherichia coli* is an another example of proteins involving MG conformation required for binding to ligands. These toxins are known to be involved in causing urinary tract infections and septicemia (Cavalieri et al. 1984). HlyA toxins are also part of RTX (repeats in toxin) family proteins, where the glycine- and aspartate-rich peptide tandem repeats are found near the C-terminal ends (Cotte 1992). HlyA is the first identified protein member of the pore-forming beta-barrel toxin family (Bhakdi and Tranum-Jensen 1991). HlyA toxins are known to form heptameric units that create a beta-barrel pore on the cell membrane after binding to specific cell surface receptors, eventually, which leads to cell death due to DNA degradation as a result of the exchange of monovalent ions through the pores (Bantel et al. 2001). Two steps are required for activation of toxins in the extracellular environment, firstly, toxins need to bind with calcium ions (Ostolaza et al. 1995; Bakás et al. 1998), and secondly, process of acylation or covalent binding to fatty acids required for oligomerization (Soloaga et al. 1996; Bakás et al. 2006; Herlax and Bakas 2007; Herlax et al. 2009). Both steps are important for binding of toxins to the membranes for their hemolytic and cytotoxic effects. Using disorder prediction analyzes, Herlax and Bakas (2007) showed that HlyA contains nine regions composed of 10–30 natively disordered amino acids. The experiments carried out using chemical denaturants, the ANS binding parameters, and protease digestions also suggested the presence of MG structure in the acylated form of HlyA toxins. To summarize the above reports, Herlax and Bakas (2007) proposed that the MG conformation of HlyA is induced by acylation process, which helps the toxins to bind to cell surface receptors such as, CD11a and CD18 subunits of β2 integrin in human target cells HL60 (Lally et al. 1997), glycoprotein glycophorin in horse erythrocytes (Cortajarena et al. 2001), and various unknown molecules during its mechanism of action.

4.3.1.4 Ribosomal Proteins

Ribosomal proteins are associated with various cellular and viral regulatory processes. Ribosomal proteins are well known to undergo folding transitions upon RNA binding required for protein-RNA recognition and in the stabilization of the resulting structure (Yonath and Franceschi 1997). Ribosomal proteins are known to exist in the disordered or MG state required for RNA binding (Fig. 4.4). Ribosomal proteins such as transcriptional antitermination N proteins of bacteriophages, λ (Su et al. 1997) and P22 (Cai et al. 1998), and the Human immunodeficiency virus type 1 (HIV-1) regulatory proteins, Tat (transactivator of transcription) (Debaisieux et al. 2012), and Rev (Karn et al. 1991) constitutes the family of proteins that possesses an RNA-binding domain with arginine-rich motif (ARM) (Bayer et al. 2005). These proteins play an important role in the regulation of HIV-1 replication and in treating latently infected cells. HIV-1 Rev and Bovine immunodeficiency virus (BIV) Tat ribosomal proteins are known to possess disordered RNA recognition regions in the solution state, which later adopt stable alpha-helical and beta-hairpin conformations upon binding to the *Rev-Response Element* (RRE) and *Trans-Activation Response* (*TAR*) recognition element, respectively (Calnan et al. 1991; Surendran et al. 2004; Shojania and O'Neil 2006; Casu et al. 2013). Another example includes the ribosomal protein L11-C76 (C-terminal domain of 50S ribosomal protein L11 in *Geobacillus stearothermophilus*), in which a 15-residue loop is known to possess a disordered and highly flexible structure that becomes structured upon RNA binding (Markus et al. 1997). L11-C76 forms a part of the ribosomal stalk that helps the ribosome to interact with GTP-bound translation factors (Markus et al. 1997). Yeast ribosomal protein YP2b is also known to display characteristics of an MG state in

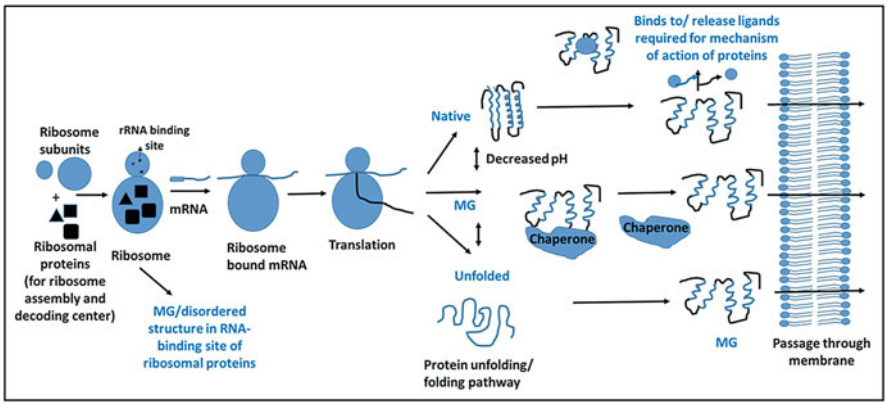

Fig. 4.4 Schematic representation of the possible roles of MG structure at different molecular levels of protein expression and cellular processes. MG state exists in the RNA-binding site of ribosomal proteins, binding or release of ligands that are required for cellular processes of proteins and during transfer to membrane surface either as MG by conversion from native or unfolded protein states due the membrane surface exposure or via the assistance of chaperones

the absence of RNA (Zurdo et al. 1997). In summary, RNA-binding sites of various ribosomal proteins undergo through the MG state transition (that makes the structure flexible) during the folding process which eventually aids in RNA-binding.

4.3.2 Protein Trafficking

The irregular shape and dynamic nature of the proteins in the MG state have been related to protein trafficking as well as translocation (Ptitsyn 1995). The close proximity of MG state proteins to the membrane aids in their easy passage through the membrane as compared to the native folded states of the protein. One explanation for the occurrence of a process of membrane transport, as suggested by Bychkova et al. (1996), is that the locally lowered pH and negative electrostatic potential around the membrane surface leads to a local region rich in proteins in their MG states. It was also proposed that membranes do not only need MG for translocation or trafficking of proteins, but the membranes also have an innate capacity to turn a protein into MG state due to the presence of local negative potential (Bychkova and Ptitsyn 1993). Eisenberg et al. (1979) proposed based on the conducted experiments and electrostatic theory, "that a membrane with a strong negative potential has the capacity to decrease the pH locally by 2 units at 5–15 Å distance from the membrane surface due to the attraction of protons". Another well known of source of decrease in local pH is the efflux of protons via the proton pumps in the membrane channels (Skulachev 1988). Schatz' group has also shown that the partial unfolding of proteins near the membrane surface occurs due to the presence of negative charges and also being similar to folding intermediates (Endo and Schatz 1988; Eilers et al. 1989; Endo et al. 1989). Along with the decrease of pH, another important feature is lower dielectric constant that strengthens electrostatic interactions near the membranes, aids in the transformation of native protein into the MG state (Bychkova and Ptitsyn 1993).

pH-induced MGs are the well-known conformational states involved in various membrane-binding proteins (Ptitsyn 1992). The precursor of OmpA protein of *E. coli* (present at the outer membrane) is the first protein for which existence of MG state near the membrane was established (Lecker et al. 1990). MG state of proteins such as cytochrome c, plastocyanin, and steroidogenic acute regulatory protein (StAR) are also known to be involved in the trafficking process across the cellular membranes, the details of the processes involved have been described below.

4.3.2.1 Cytochrome c

Cytochrome c, a heme protein, is involved in the electron transport chain at the inner membrane of mitochondria. It is one of the examples of MG state proteins involved in protein trafficking. Occurrence of a MG state has been established in cytochrome

c by two methods: (1) using low pH and high salt concentration (pH 2.0 and 0.5 M KCl) conditions (Ohgushi and Wada 1983), and (2) methanol-induced conformational transitions (moderately low pH and moderately low dielectric constant) which simulates the physiological conditions near the negatively charged membrane surfaces (Bychkova et al. 1996). The MG state showed the conservation of native secondary structure with loosely packed protein interior. The MG state of cytochrome c has been proposed to be involved in an intermediate state of protein folding which could be important for its function in the cells. Various spectroscopic techniques such as near- and far-UV CD, Trp fluorescence, microcalorimetry, and diffusion measurements have been used to establish the occurrence of the MG state in cytochrome c (Bychkova et al. 1996). De Jongh et al. (1992) also demonstrated the existence of the MG state for cytochrome c in the presence of 12-PN/12-Pglycol (9,1) micelles using NMR. The MG state was found to partially lose the tertiary structure, which was remarkably different from its aqueous solution structure of tightly folded protein with secondary and tertiary structure levels. Apocytochrome c, which is a precursor of cytochrome c was also suggested to exhibit secondary folding changes upon interaction with micelles in comparison to its random coil structure in solution phase (de Jongh et al. 1992). The authors also proposed that this lipid-induced dynamic folding or MG state of the cytochrome c structure could be possibly responsible for the translocation of proteins across the membranes.

4.3.2.2 Plastocyanin

The plastocyanin is a photosynthetic copper-containing monomeric protein (present in green plants and many algae), involved in electron transfer in photosynthesis at the lumen-side of the thylakoid membrane of chloroplasts. It remains unfolded as a high molecular mass precursor protein (apo-plastocyanin) before transport into the chloroplast (Merchant and Bogorad 1986; Li et al. 1990) and is known to refold very slowly *in vitro* (Koide et al. 1993). The NMR experiments of french bean apo-plastocyanin by Koide et al. (1993) suggests the presence of compact intermediate state with flexible beta-sheets and altered the packing of the hydrophobic core during the refolding reaction, indicating the presence of MG state.

4.3.2.3 StAR

StAR is known for its involvement in the transport of cholesterol from the outer leaflet to the inner leaflet of mitochondrial membrane. The active form of StAR has been shown to exist in a MG state (Bose et al. 1999) that facilitates mitochondrial entry of the protein as well as acts as a switch control for cholesterol entry into the mitochondria. In this case protein, proximity to the membrane has also been postulated to act as a determining factor for the formation of MG state.

4.3.3 Role of MG States in Chaperone Function

Chaperones are well known to interact with partially denatured or folded intermediates of proteins. Basically, chaperones prevent the proteins from folding too fast as well as prevent them aggregating. Remarkably, chaperones have also been associated with the transfer of MG state proteins across the cell membranes or for involvement in the biological processes (Fig. 4.4). Some of the well-known examples of the chaperones that binds to MG state proteins include GroEL, alpha-crystallin, Domain V of 23S rRNA, and peptidylprolyl cis-trans-isomerase (PPI).

The chaperone GroEL is known for its binding to MG state intermediates which expose hydrophobic surfaces during early stages of folding (Laminet et al. 1990; Martin et al. 1991; Weissman et al. 1995). The alpha-crystallin is also known for its ability to interact with aggregation-prone refolding intermediate of lysozyme as well as for the formation of stable complexes with the MG state of alpha-lactalbumin and bovine carbonic anhydrase (BCA) (Raman et al. 1997; Rajaraman et al. 1996, 1998). BCAII and lysozyme are known to form equilibrium MGs under mildly denaturing conditions (Ikeguchi et al. 1986; Uversky et al. 1992; Dobson et al. 1994; Radford et al. 1992; Wang et al. 2010). Domain V of 23S rRNA (acts as chaperone) of *E. coli* ribosome has been demonstrated to be able to interact with various folded states of BCAII and chicken egg white lysozyme MG structure, and it was also found to be suppressing the aggregation and reactivation states under refolding conditions (Pathak et al. 2014). The results were also supported by the experiments conducted with variants of domain V RNA and its mutants where the retaining time of chaperone bound substrate state played an important factor in determining the ability of chaperones to prevent or suppress aggregation (Pathak et al. 2014). PPI has also been shown to bind to it's substrate, MG state BCA which helps in prevention of aggregation and eventually helps in fastening of the last steps of folding (Freskgård et al. 1992). The prevention of aggregation was found to be related to restoration of 100 % enzymatic activity in PPI (Freskgård et al. 1992). PPI has also been reported to be identical to the cyclophilins, a class of heat-shock proteins (molecular weight 56–59 kDa) (Schreiber 1991). The above-mentioned reports suggested that the PPI might be acting as a chaperone in initial stages of folding for binding with MG state substrate and acts as a catalyst during the last steps of folding.

4.4 Impact of MG States on Enzymatic Activity

MG states of most of the enzymes are known to be inactive in terms of their enzymatic activity (Ptitsyn et al. 1990; Tsou 1995). However, MGs of enzymes such as xylanase and dihydrofolate reductase have been shown to retain some enzymatic activity (López et al. 2011; Uversky et al. 1996; Vamvaca et al. 2004), or no

activity in case of homocysteine thiolactone (Kumar et al. 2014b). Other than the above-mentioned enzymes, endopeptidase activity of botulinum neurotoxin (BoNT) or its catalytic domain have been demonstrated to retain full enzymatic activity in the MG state (Cai and Singh 2001; Kukreja and Singh 2005; Kumar et al. 2014a).

4.4.1 Botulinum Neurotoxins

BoNTs are the most potent toxins reported to date (Gill 1982; Arnon et al. 2001). The BoNT/A structure consists of two major domains, heavy chain (HC, 100 kDa) and light chain or catalytic domain (LC, 50 kDa) connected via a disulfide bond and non-covalent interactions. The cleavage of disulfide bond takes place in the cytosol and is vital for its toxic effects. HC is involved in the binding to neurons and subsequent formation of channel across the endosomal membranes, and LC is involved in the cleavage of synaptosomal-associated protein, 25 kDa (SNAP-25) at the presynaptic membrane, thus inhibiting neurotransmitter (acetylcholine) release (Lacy et al. 1998; Schiavo et al. 2000; Singh 2000) and causing flaccid muscle paralysis. BoNT is the only known protein which has a fully enzymatically active MG state that makes it a biologically important MG structure (Cai and Singh 2001; Kukreja and Singh 2005). Cai and Singh (2001) for the first time reported the presence of a fully enzymatically active MG state in disulfide reduced form of BoNT type A, under physiological conditions, i.e. 37 °C. The CD spectroscopy and ANS binding results showed the presence of significant native secondary structure along with the loss of most of the tertiary structure. The temperature profile of BoNT type A showed a strong correlation with the existence of the MG structure and optimum endopeptidase activity at 37 °C under reducing conditions. Remarkably, the non-reduced BoNT/A showed enzymatically inactive toxin with no MG structure, indicating the importance of reduction of the disulfide bond in the occurrence of enzymatic activity of the LC that leads to its toxicity. Kukreja and Singh (2005) also reported the presence of a prominently biologically active 61 % molten-globule structure of catalytic domain of BoNT/A, i.e. LC at 50 °C. Notably, the same study also reported the presence of an novel pre-imminent molten-globule enzyme (PRIME) state of BoNT/A catalytic domain that is optimally active. Under the PRIME conformation, the catalytic domain retained its secondary structure along with fluidic tertiary structure, which is more likely to accommodate specific interaction with the substrate, SNAP-25, ultimately aiding in selective and optimum enzymatic activity (Kukreja and Singh 2005). The finding of active MG state in BoNT is of utmost importance in view of it being the most toxic protein as well as regarding its extreme specificity to the substrate, which is important for it a mechanism of action and cellular toxicity. The MG state of BoNT will not only help in unfolding of the mechanism of action but will also provide vital information on enzymatically active molten-globule structures.

4.4.2 Homocysteine Thiolactone

Another example of a process involving the role of MG state in enzyme activity is of homocysteine thiolactone. Homocysteine thiolactone is produced from homocysteine via a process known as 'protein N-homocysteinylation', resulting in the formation of amide bonds with a ε-amino group of lysine residues. Homocysteine is a byproduct of methionine metabolism pathway. The production of homocysteine thiolactone results in the loss of enzyme function, generation of oxidative stress and occurrence of an autoimmune response against N-homocysteinylated self-protein (Undas et al. 2004; Kumar et al. 2014a, b). Protein N-homocysteinylation has also been reported for formation of toxic aggregates or amyloid formation, both of which are involved in neurodegenerative diseases (Mattson and Shea 2003; Obeid and Herrmann 2006; Paoli et al. 2010; Stroylova et al. 2012). Studies on native proteins that undergo covalent modification by targeting lysine residues have shown the existence of MG state leading to aggregate formation. Kumar et al. (2014a, b) reported the effect of protein N-homocysteinylation on the structural conformation of three different proteins (cytochrome-c from bovine heart, bovine alpha-lactalbumin and lysozyme from chicken egg white) at different time intervals. All three proteins differ in physicochemical properties and lysine residues. The results suggested that the protein N-homocysteinylation results in the formation of MG state in cytochrome-c and alpha-lactalbumin, however, no structural changes were observed in lysozyme. The authors proposed that the MG state might act as an intermediate during the protein folding pathway, which could be responsible for protein aggregation and eventual loss of enzyme activity.

4.5 Translocation of Toxins/Proteins Across the Membranes

MGs acting as partly unfolded denaturation states have been proposed to be an essential requirement for the translocation of proteins or toxins across the membranes (Bychkova et al. 1988) either by insertion, penetration, or channel formation. Some toxins, as well as proteins, are known to undergo through MG intermediate states during membrane insertion at low pH conditions required for their biological activity. Bychkova et al. (1988) points out that the translocation competent state could be MG state owing to the four properties of protein at the membrane surface, firstly, non-native structure of protein at membrane required for translocation; secondly, existence of this non-native structure as an folding intermediate; thirdly, high potential barrier between native and non-native structure required for translocation; and lastly, existence of this state for a long period after renaturation. Eliers et al. (1988) introduced a very important aspect of the translocation of proteins across the membranes by stating that "a cell organelle can partly unfold on its surface precursors which must be imported into this organelle and that it is unlikely that this reaction is physiologically irrelevant". The concept of MG as a translocation competent

state started from the above-mentioned reports. Many toxins have also been reported to be involving MG state for their translocation across the host cell membrane to exert their functions or toxic effects.

Toxins are substances created by plants and animals that are poisonous to humans. Toxins at low doses can also act as drugs, but turns out to be poisonous once used in large amounts. Apart from the biological functions, they also have an interesting structural organization. Various toxins have been reported to exist in MG conformation at different stages of their biological function. For example, the acid-induced transition leads to the formation of MG state from the native state during the endocytosis step. Some of the toxins that undergo acid induced transitions include colicins (Bourdineaud et al. 1990a, b; van der Goot et al. 1992; Lakey et al. 1992), diphtheria toxin (DT) (Zhao and London 1986; Ren et al. 1999), *Pseudomonas* exotoxin A (Farahbakhsh and Wisnieski 1989; Jiang and London 1990), *Staphylococcus aureus* α-toxin (Vécsey-Semjén et al. 1996), botulinum neurotoxin (BoNT) (Cai and Singh 2001; Cai et al. 2006), and anthrax toxins (Krantz et al. 2004).

4.5.1 Colicins

Colicins were the first toxins where the correlation between MG structure and membrane insertion or translocation competent state was established (Bourdineaud et al. 1990a, b; van der Goot et al. 1992; Lakey et al. 1992). Colicin toxins are a group of toxins known to kill cells via loss of potassium ions. The colicin toxin A constitutes three major domains each responsible for one of the three steps involved in its mechanism of action. The mechanism of action for colicin includes three major steps—(1) receptor binding of colicin on the outer membrane of the cell, (2) translocation through the membrane, and (3) permeabilization or pore formation of the membrane. The C-terminal domain is responsible for pore formation and its consists of ten α-helices out of which eight are amphiphilic which surrounds the nonpolar hairpin formed by the remaining two helices, also protecting the core from water (Parker et al. 1989). The toxin penetration into the membrane was found to be dependent on the transition of the C-terminal or pore-forming domain of colicin A into the MG state as a result of a decrease in pH near the negatively charged membrane (van der Goot et al. 1991). The pore formation was found to increase between pH 5.5 and 4.2 (van der Goot et al. 1991), suggesting a role of low pH in the translocation process. The positively charged phospholipids were proposed to either orient the molecule at the bilayer surface (Parker et al. 1989) or decrease the surface pH locally for facilitation of formation of MG structure (González-Mañas et al. 1992). To sum up, van der Goot et al. (1992) suggested two explanations for the existence of MG state in colicin A, (1) the local pH near the membrane surface leads to the denaturation of the pore forming or C-terminal domain, that transforms the native structure into MG state (Lakey et al. 1992). The MG state converts the pore forming domain into the more flexible structure, eventually allowing the

hydrophobic hairpin to form pores in the membrane. (2) Formation of MG state by either binding to the receptor proteins or chaperones during the translocation step.

Colicin E1 (Merrill et al. 1990), another type of colicin toxins shows a similar translocation mechanism to colicin A toxin. However, colicin N, which shares >50 % sequence identity with colicin A, does not show the formation of an MG state at low pH conditions unlike colicin A (Fridd and Lakey 2002). Notably, it is possible for the subtypes of the same toxin to follow different translocation mechanisms. The δ toxin from *Bacillus thuringiensis* (Li et al. 1991) has also been reported to involve a MG state at low pH for pore formation. The above-mentioned reports of involvement of a MG state in pore formation provided a new direction to the understanding of the translocation process of toxins.

4.5.2 Staphylococcus aureus *α-Toxin*

Staphylococcus aureus α-toxin is an another example of the toxins that utilizes a MG state for channel or pore formation. It was the first bacterial exotoxin known to be a pore former. *Staphylococcus aureus* secretes a pore forming α-toxin as a soluble monomeric protein. These toxins act on rabbit erythrocytes, human platelets, monocytes, endothelial cells, and non-specific receptors. At high concentration, this toxin causes membrane damage as well as many secondary cellular reactions triggered via a calcium ions influx through the pores formed (Bhakdi and Tranum-Jensen 1991). Few of such reactions include the stimulation of arachidonic acid metabolism, triggering of granule exocytosis, and contractile dysfunction (Bhakdi and Tranum-Jensen 1991). The C-terminal half of the α-toxin has been known to be required for channel formation. Various spectroscopic techniques have been used to analyze the effect of low pH on the kinetics of channel formation as well as on the conformation of the toxin, suggesting the occurrence of protein unfolding (Vécsey-Semjén et al. 1996). However, the most important step in unfolding was found to be related to the C-terminal half of the toxin, which is of utmost importance in consideration to MG state. The MG state has been suggested to be occurring simultaneously with the exposure of hydrophobic regions on the surface of the protein as well as with the onset of channel formation (Vécsey-Semjén et al. 1996). Overall, the results by Vécsey-Semjén et al. (1996) suggested that the C-terminal half of the toxin needs to undergo partial unfolding or MG state formation for channel formation.

4.5.3 Diphtheria Toxin (DT)

DT includes two major fragments, namely A (27 kDa) and B (37 kDa), linked via a disulfide bond (London 1992a, b, Choe et al. 1992). Fragment A is a catalytic domain which gets released into the cytosol and kills the cell by ribosylating

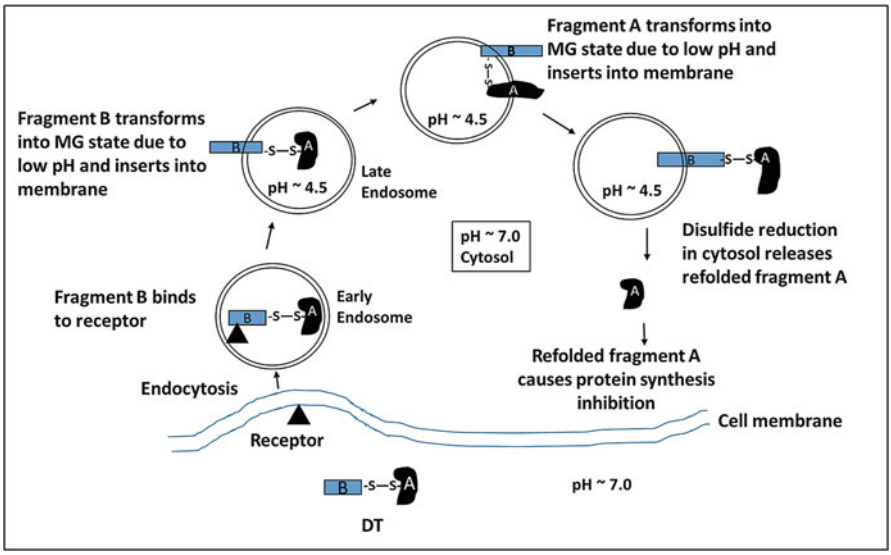

Fig. 4.5 Schematic representation of the role of MG state in DT intoxication. DT binds to the receptor and undergoes endocytosis at the cell membrane. In the endosomes, first fragment B and later fragment A undergo acid-induced transition to the MG state that aids in their insertion into the endosomal membrane. Fragment A then passes through the pore and enters the cytosol where its disulfide bond is reduced due to reductases. Fragment A refolds at cytosol pH 7.0 and causes inhibition of protein synthesis

elongation factor 2, which is required for protein biosynthesis. Fragment B is further divided into two domains—a receptor-binding domain which consists of a flattened β barrel, and translocation domain consists of nine helices, with two hydrophobic alpha-helices, buried inside (Choe et al. 1992). The translocation or pore-forming domain of DT is similar to colicin A (Parker et al. 1989) and δ toxin (Li et al. 1991). DT have been studied extensively in terms of existence of MG state in both fragments A and B at low pH as well as higher temperature conditions (Dumont and Richards 1988; Zhao and London 1988; Cabiaux et al. 1989; Ramsay et al. 1989; Ramsay and Freire 1990; Jiang et al. 1991; London 1992a, b; van der Goot et al. 1992; D'Silva and Lala 1998). The MG state of DT has also been demonstrated to correlate with the degree of membrane insertion (Papini et al. 1987; Dumont and Richards 1988; Zhao and London 1988; Cabiaux et al. 1989). van der Goot et al. (1992), suggested the coexistence of DT MG state with the events of insertion into the endosomal membrane, and translocation process (Fig. 4.5) as following, firstly, fragment B of DT transforms into a MG structure which aids in its insertion into the endosomal membrane for channel formation. Then, the fragment A undergoes transformation into MG state at low pH of the endosome, which leads to its insertion as well as translocation across the membrane. Remarkably, the MG state of fragment A is reversible as compared to fragment B (which is irreversible) indicating refolding of fragment A back into the native structure in the cytoplasm (pH 7.0) to exert its activity or function.

4.5.4 Pseudomonas aeruginosa *Exotoxin A*

The translocation, or pore forming, domain of *Pseudomonas aeruginosa* exotoxin A is majorly alpha-helical (Allured et al. 1986) and similar to the DT in the overall organization of three domains. These toxins also undergo endocytosis in the cells to blocks protein synthesis by ADP-ribosylation of elongation factor 2, causing cell death. These toxins have been suggested to undergo a partial unfolding or MG state existence in order for insert into the membrane (Farahbakhsh and Wisnieski 1989; Jiang and London 1990; Ptitsyn 1995).

4.5.5 *Role of MG State in Channel Formation for BoNT*

The MG conformation of BoNT LC type A (BoNT/A LC) had been proposed to have a role in translocation of BoNT/A LC across the endosomes to the cytosol (Cai and Singh 2001; Cai et al. 2006). The reduced form of BoNT/A LC has been known to adopt MG state after undergoing unique structural changes under low pH conditions, which exposes a substantial number of hydrophobic groups (Cai et al. 2006). The MG state is associated with the retention of secondary structure, and significant loss of tertiary structure as determined by using circular dichroism (CD) spectroscopy and enhancement in ANS fluorescence intensity upon its binding to LC at low pH conditions (Cai et al. 2006) (Figs. 4.6 and 4.7). The tryptophan fluorescence

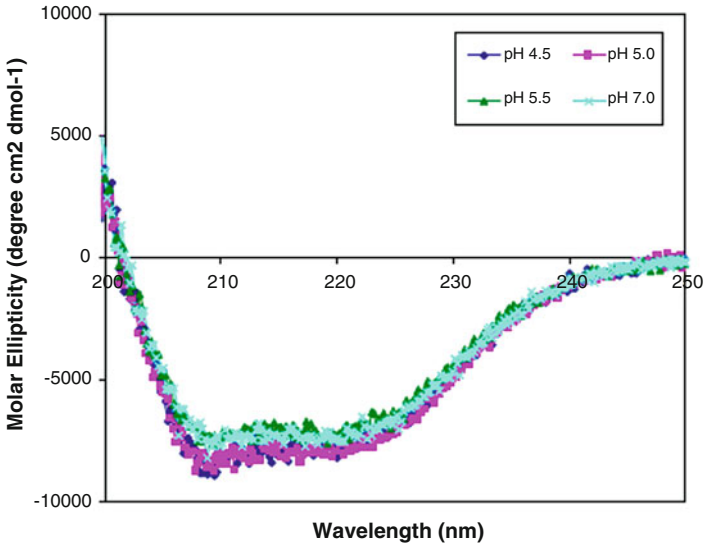

Fig. 4.6 Far UV CD spectra of BoNT/A LC at pH 7.0, 5.5, 5.0, and 4.5. The UV spectra showed no significant changes in the molar ellipticity at 222 nm for pH 5.5, 5.0, and 4.5 as compared to pH 7.0. The results indicate no significant changes in the secondary structure of BoNT/A LC with pH change (Reproduced with permission, Cai et al. 2006)

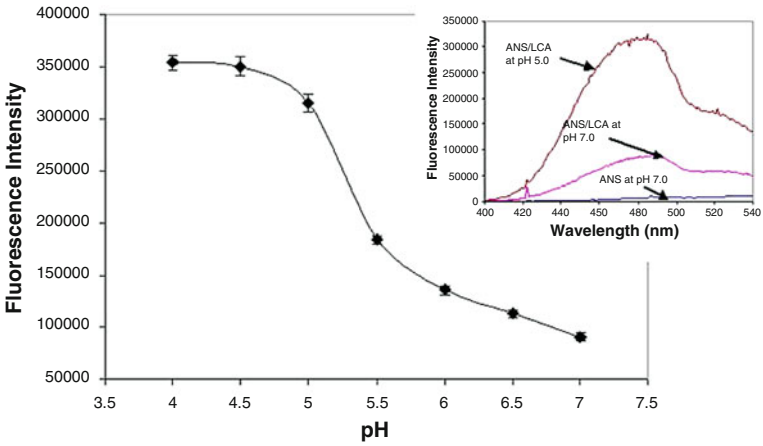

Fig. 4.7 Effect of ANS binding of BoNT/A LC on the fluorescence intensity as a function of pH. The fluorescence intensity was recorded at 488 nm with excitation wavelength of 370 nm. The insert figure shows fluorescence emission spectra of ANS and ANS bound BoNT/A LC at pH 7.0 and 5.0. The error bars represent the standard deviation from three independent experiments. The ANS fluorescence intensity increases for ANS bound BoNT/A LC with decrease in pH with significant changes at pH 5.0–5.5. The exposure to acidic pH results in conformational changes in the protein due to exposure of LC hydrophobic surfaces or clusters to ANS, suggesting a major structural transition (Reproduced with permission, Cai et al. 2006)

lifetime distribution of BoNT/A LC also showed a significant broader peak at pH 5.0, suggesting a significant increase in the heterogeneity of emission at low pH (Cai et al. 2006). The MG state in BoNT/A LC has been proposed by Cai et al. (2006) to be required for translocation of BoNT/A through the narrow endosomal membrane channel, which is approx. 15 Å observed in electron micrographs of BoNT/B in two-dimensional lipid arrays (Schmidt et al. 1993). The proposed mechanism of Cai et al. (2006) involves tertiary structural level changes in BoNT/A LC at low pH which leads to formation of an elongated MG molecule, however retaining specific interactions with the HC and lipid bilayer through hydrophobic clusters which are critical for passage of LC through the HC channel under a refolded state (Fig. 4.8). Similar to DT, BoNT/A LC also further renatures or refolds in the cytosol, and regains its endopeptidase activity. Hence, BoNT/A presents a unique example of molten globule state involved in channel formation at low pH conditions, being temperature independent.

4.5.6 Anthrax Toxin

Anthrax toxin consists of three major components, the translocase component, protective antigen (PA), and two enzyme components: lethal factor (LF, a Zn^{2+} metalloprotease) and edema factor (EF, a calcium ion, and calmodulin-activated

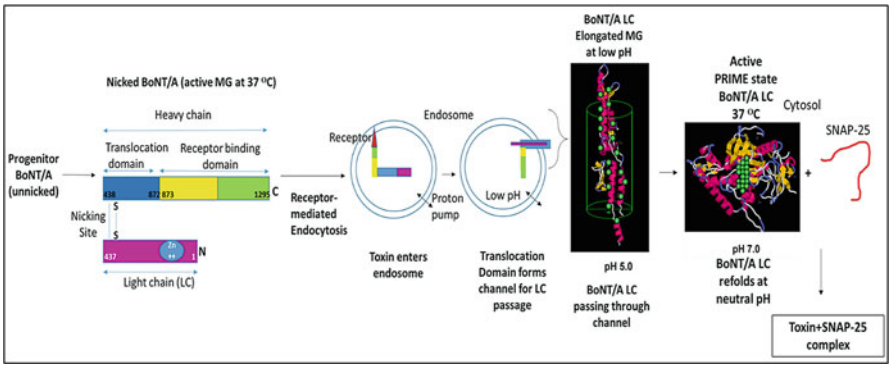

Fig. 4.8 Schematic representation of MG state existence in Botulinum neurotoxin type A. The green colored balls represent the non-polar side chains which are tightly packed in the native conformation of BoNT/A at pH 7.0. At pH 5.0, BoNT/A LC undergoes transformation into an elongated MG state structure, wherein the hydrophobic regions are exposed to the surface of the channel retaining specific interactions with the HC and lipid bilayer, eventually aiding in its translocation through the HC channel. The PRIME state facilitates the binding of substrate SNAP-25 at 37 °C (Reproduced with permission, Cai et al. 2006)

adenyl cyclase) (Collier and Young 2003). Anthrax toxins are known to enter mammalian host cells via acidified endocytic route and once it's inside the cytosol the enzymatic components disrupt cell-signaling pathways which lead to general immune system dysfunction and eventually death. Active holotoxin complex consists of a ring-shaped PA oligomeric (octameric or heptameric) pre-channel which can bind to 3/4 copies of LF and EF. Under the acidic conditions of endosomes, the PA pre-channel undergoes a conformational change required for insertion into the membrane by forming a cation-selective channel. The proton motive force and low pH in the endosomes unfolds and translocates enzymatic components LF and EF through the protein-conducting channel formed by translocase component PA (Krantz et al. 2005; Thoren et al. 2009). The solution state studies of EF and LF indicate toward the destabilization of under mildly acidic pH conditions (pH 5–6) (Krantz et al. 2004). It was also shown in the same study that N-terminal of LF and EF shows the formation of compact MG-state under mildly acidic pH conditions, displaying major secondary structure and disrupted tertiary structure (Fig. 4.9). Krantz et al. (2004) also suggested that the toxins with a large amount of secondary structure can be translocated through the narrow membrane channels, however, same cannot be attained with rigid tertiary structure, indicating towards transformation into the MG state. An interesting model proposed by Krantz et al. (2005, 2006) gives a novel approach to the mechanism about how the MG state of a protein (partially unfolded) could be fully unfolded and linearized during translocation (Fig. 4.8). Krantz et al. (2005) identified Phe-427, a hydrophobic residue at the channel's cap region, and showed the presence of the small hydrophilic residues lined up around the channel lumen along the cap as well as the β-barrel stem region (Nassi et al. 2002). Their studies further indicated the presence of a

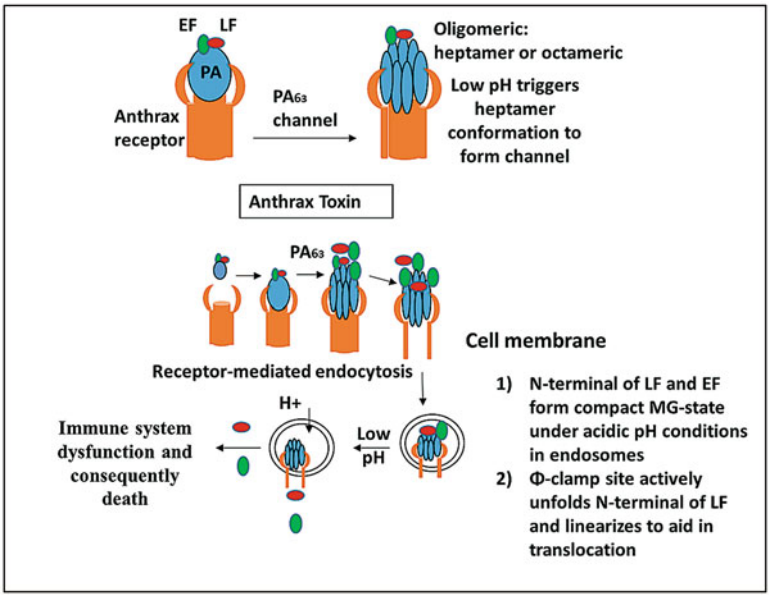

Fig. 4.9 Mechanism of action and illustration of role of MG state of anthrax toxin. *PA* denotes protective antigen, *LF*—lethal factor, and *EF*—edema factor. Active holotoxin translocase complex consists of a *ring-shaped* PA oligomeric pre-channel which can bind to 3/4 copies of LF and EF. Under the acidic conditions of endosomes, the PA pre-channel undergoes a conformational change and inserts into the membrane. Low pH and proton gradient induces MG state in the N-terminal of LF and EF structure, inside the endosomes to aid in their translocation into the cytosol. Φ-clamp (Phe residues constitute hydrophobic core) leads to complete unfolding of the substrate from the MG state during translocation, thereby reducing the energy penalty of exposing hydrophobic sequences

radially-symmetrical aromatic clamp site (made up of Phe residues) known as "Φ clamp" (Krantz et al. 2005), which is required for translocation of the protein, by acting as the polypeptide binding site in the channel. Mutation studies of these residues suggested inefficient translocation of the protein (Krantz et al. 2005; Janowiak et al. 2010). Further studies by Thoren et al. (2009) also suggests that the Φ-clamp site actively unfolds N-terminal of LF during translocation. The studies suggest the role of binding of the translocation chain to the Φ-clamp (hydrophobic core) for the efficient unfolding of the substrate during translocation (Fig. 4.9). The clamp region creates an surrogate environment similar to the hydrophobic core of the unfolded protein that helps in complete substrate unfolding from the MG state, thereby reducing the energy penalty of exposing hydrophobic sequences to solvent (Krantz et al. 2006). Krantz et al. (2004) showed that the substrate unfolding and translocation occurs by stabilization of the unfolding intermediates via a series of interactions with many chemical groups and α-helical structure. Krantz et al. (2006) suggested that the cooperative functioning of ΔpH driving force and a functioning Φ-clamp site is required for the unfolding of the substrate for efficient translocation process (Thoren and Krantz 2011).

4.5.7 Role of Pre-MG or MG State in RTX Toxins

Some proteins exhibit partial folding of the intrinsically unfolded proteins in the presence of cations such as calcium, and zinc. Example of calcium binding proteins includes Repeat-in-toxin (RTX) toxins and that of zinc binding proteins includes, human sperm protamines (P2 and P3) (Gatewood et al. 1990), thymosin α 1 (Grottesi et al. 1998), prothymosin (Uversky et al. 2000b), and phosphodiesterase γ-subunit (Uversky et al. 2002). Uversky et al. (2001) demonstrated partially folded Human α-synuclein in the presence of several divalent and trivalent metal ions. These proteins have been shown to be transformed into the pre-MG state from intrinsic coils. The RTX toxins require a pre-MG or MG state domain for passage through the channel and dependent on the calcium ion presence instead of low pH as observed in other toxins such as colicin, DT, anthrax, and BoNT. RTX proteins are the group of cytotoxins which are produced by a wide range of gram-negative bacteria (genus including *Bordetella*, *Escherichia*, *Proteus*, and *Pasteurellaceae*) (Lally et al. 1999; Frey 2011) as well as by *Vibrio* genus (*V. cholerae* and *V. vulnificus*, produces mainly multifunctional autoprocessing RTX toxins (MARTX)) (Linhartová et al. 2010). RTX motifs commonly occur as C-terminal located glycine and aspartate-rich sequences of nine amino acids referred to as [GGXGXDX[L/I/V/W/Y/F]X] motif (where X represents any amino acid) (Satchell 2007). These proteins are also known as toxins because of various secreted lipases and proteases involving the RTX motif. The RTX toxins can be divided into two groups, pore-forming leukotoxins (including hemolysins) and MARTX toxins. RTX family toxins are mainly characterized by two major features, (1) number of repeats in the toxin protein sequences, and (2) ATP-binding cassette (ABC) transporter-based system for extracellular secretion by the type I secretion system (T1SS). T1SS proteins are known to form oligomeric secretion channels spanning both the outer and inner membrane of the bacterial cell wall. T1SS recognizes first the C-terminal motif of RTX toxins, hence aiding in the passage of these toxins via channels from bacterial cytosol directly into the extracellular space by preventing their exposure to the periplasmic space between the outer and inner membrane (Linhartová et al. 2010). The RTX motifs act as a site of calcium binding, which helps in RTX protein folding following its export via T1SS, the latter being encoded within the *rtx* operon (Linhartová et al. 2010). The RTX cytotoxins share a common feature of cation-selective pore formation among their toxin family (Bumba et al. 2010). RTX toxins/proteins display a diverse number of functions owing to their belonging to a large and diverse family of proteins. Some of the functions include *Cyanobacteria* motility, acting as synergistic virulence factor that causes tissue damage by stimulating the production of inflammatory mediators, acting as bactericins, rounding of epithelial cells occurring due to catalyzed covalent cross-linking of actin (*Vibrio cholera*), and many others. However, the biological role of RTX toxins is still not well understood.

The calcium ion requirement in RTX toxins was first reported for Uropathogenic *Escherichia coli* α-hemolysin (HlyA) (Short and Kurtz 1971) and *Bordetella pertussis* adenylate cyclase toxin (CyaA) (Hanski and Farfel 1985). RTX proteins are well known to possess hallmarks of intrinsically disordered proteins in the absence

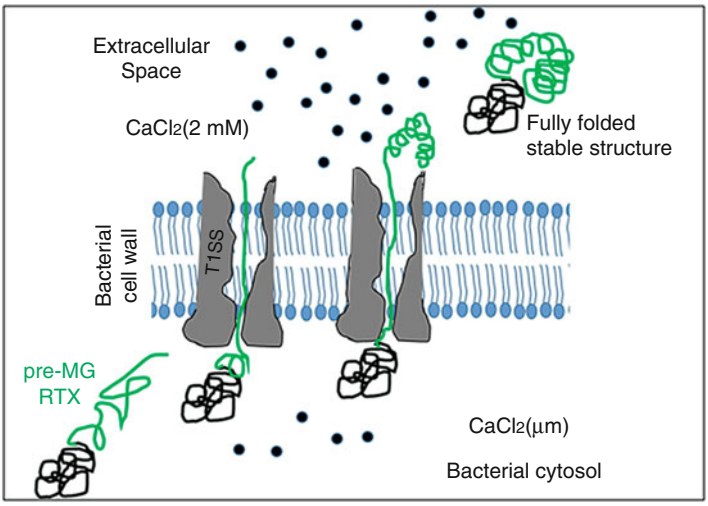

Fig. 4.10 Diagrammatic representation of the involvement of pre-MG state in the process of secretion of RTX toxins/proteins, dependent in calcium ion concentration. Calcium ion are represented by *black circles*. RTX domain is highlighted in *green color* and RTX protein (CyaA) in *black color*. Pre-MB state of RTX protein is shown to pass through the T1SS pore to the extracellular space (Open Access, Sotomayor-Pérez et al. 2015)

of calcium. The pre-MG state of RTX proteins has been proposed for aid in their secretion through the channel formed by T1SS (Fig. 4.10). In general, the disordered structure leads to the adoption of pre-MG state and displays a strong hydration which occurs due to the electrostatic repulsions between negatively charged residues in the structure (Sotomayor-Pérez et al. 2015). The disordered or pre-MG state has been proposed to be induced in RTX due to either of the following, electrostatics binding to a cell-surface receptor, pH or membrane association (Sotomayor-Pérez et al. 2015). The calcium binding (2 mM $CaCl_2$) process leads to the folding and subsequent stabilization of the secondary as well as tertiary structure states of RTX by a reduction in the net charge, dehydration, and compaction of the structure (Fig. 4.10). Hence, protein secretion is regulated by calcium gradient across the bacterial cell walls via pre-MG state formation. Two examples of the RTX toxins showing pre-MG state, namely, CyaA and HlyA have been discussed below.

CyaA is one of the major virulence factors of *Bordetella pertussis* (known to cause whooping cough) (Weiss and Hewlett 1986; Ladant and Ullmann 1999). The C-terminal region of receptor binding domain (701 amino acids) of CyaA contains ~40 calcium-binding RTX motifs (Cotte 1992; Welch 2001). Chenal et al. (2009) showed the existence of disordered, highly hydrated and weak structure of RTX domain in the absence of calcium. In the same report, authors showed compaction and dehydration of the RTX domain upon addition of calcium ions leading to

the formation of stable 3-dimensional structure. Chenal et al. (2010) showed the existence of a pre-MG state in this toxin in the absence of calcium using various spectroscopic techniques (far-UV CD, Trp fluorescence, Fourier transform infrared spectroscopy (FTIR)). The pre-MG state of RTX has been characterized by high content of random coil structure and a lower amount of β-sheets and turns (Chenal et al. 2010). The thermodynamic analysis was performed by using urea or guanidine chloride induced protein denaturation (Chenal et al. 2010). To sum up, the pre-MG conformation change occurs due to the strong internal electrostatic repulsions in the RTX motif between the negative charges of the aspartate-rich polypeptide sequence. On the other hand, calcium binding showed strong compaction as well as the presence of secondary and tertiary structures indicating a functional 3-Dimensional structure. The sodium ions were also shown to be able to screen the charges, however, allowing only partial compaction of the protein without inducing any folding which leads to the formation of a 3-Dimensional structure. The role of pre-MG state exists in the transfer/secretion of CyaA from the bacterial cytosol (a low-calcium environment) to external medium via a channel formed by T1SS (Chenal et al. 2010) (Fig. 4.10). Once in the external medium, CyaA adopts its functional structure in the presence of calcium ions i.e. turning into an active, receptor-binding competent conformation (Fig. 4.10).

Similarly, HlyA has also been proposed to exhibit MG state required for its function. HlyA is known to be cytolytic towards heme-rich erythrocytes and also other cell types such as epithelial cells and leukocytes from human and animal hosts (Cavalieri et al. 1984; Forestier and Welch 1991). Similar to other RTX toxins, HlyA (110-kDa multidomain polypeptide), the C-terminus region constitutes hallmark RTX domain with a consensus sequence, GGXGXDXψX (ψ represents a bulky hydrophobic residue) (Welch 1991, 2001). This domain once bound to calcium has been shown to possess a β-roll structure, which stabilizes the structure due to conformational changes exposing peptide sequences/surfaces required for interaction with host cell membranes (Sánchez-Magraner et al. 2007, 2010, Schindel et al. 2001). Wiles and Mulvey (2013) proposed a mechanism for HlyA pore formation based on the earlier reports as following, The pore formation starts as a result of bound calcium to HlyA monomers, which leads to formation of electrostatic interactions within N and C-termini resulting in binding to specific surface receptors on the target cells (Lally et al. 1999; Ostolaza et al. 1997; Bakás et al. 1996). Once associated with the target cell membrane, the MG state of HlyA toxins interacts with the outer leaflet proceeding toward the membrane insertion process involving irreversible anchoring of the domain (Soloaga et al. 1999; Herlax and Bakas 2007). Herlax et al. (2009) showed oligomerization of HlyA within host cell membranes by using fluorescence resonance energy transfer (FRET). The protein-protein interactions at the membrane are suggested to be the result of exposure of disordered regions of the HlyA toxins (Herlax et al. 2009). As a result, HlyA forms oligomeric pores in the host cell membranes depending on it lipid raft content (Herlax et al. 2009; Bakás et al. 2006) and exerts its function.

4.5.8 *Proteins Involving MG State in Translocation Process Independent of pH*

Other than the toxins many proteins or bacteriophages also utilize MG state for their insertion into the cell membrane independent of acidic pH. Few examples, include fd bacteriophage, human complement C9 protein, and myelin basic protein (MBP). Filaments of fd phage are known to convert into shortened rods known as I-forms at low temperature and then to spheroidal structures known as S-forms at a slightly higher temperature, demonstrated by chloroform–water interface experiment (Griffith et al. 1981; Manning et al. 1981, 1982; Manning and Griffith 1985; Dunker et al. 1991a). I-forms have been proposed to mimic the coat proteins that insert into the inner membrane of the *E. coli* (host) and S-forms as the membrane inserted forms (Dunker et al. 1991a; Roberts and Dunker 1993). The I- and S-forms are rich in α-helix structure and are known to exist in a MG state (exhibiting non-rigid side-chain packing) that aids in membrane penetration of fd phage into the host membrane. The four major observations that support the existence of MG state are as following: (1) both I- and S- forms exhibit substantial fragility when subjected to conditions for electron microscopy (Dunker et al. 1991b) (Fig. 4.11), (2) tend to aggregate with increasing temperature (Dunker et al. 1991b), (3) showed sensitivity of tryptophan fluorescence to nonpolar quenchers suggesting their absorbance into the hydrophobic interiors of the I-forms and S-forms due to non-rigid side-chain packing (Roberts and Dunker 1993), and (4) enhancement of extrinsic fluorescence of ANS (Dunker et al. 1991b). The non-rigid side-chain packing of the intermediate forms helps in sliding of subunits of fd phage which aids in the phage infection process (Dunker et al. 1991a).

Human complement C9 protein has been suggested to have MG state involvement for membrane penetration. This protein has been shown to exist in a partially denatured form at 36 °C (Lohner and Esser 1991), which coexists with the process

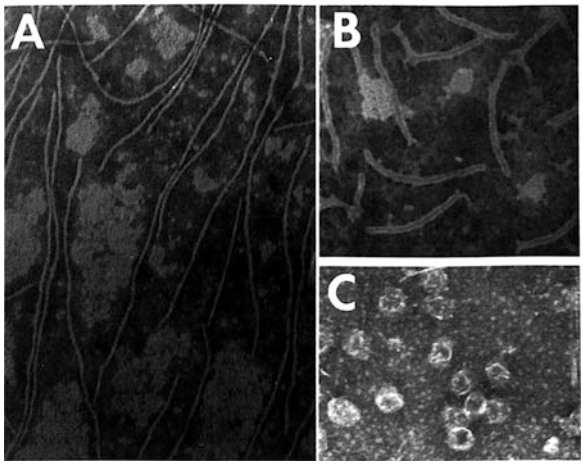

Fig. 4.11 Electron micrographs of fd phage, I-form and S-forms treated using chloroform/water two phase system. Panel (**a**) fd phage, Panel (**b**) I-form, Panel (**c**) S-form. The fd phage panel shows filamentous structures, the I-form shows the shortened rod structures, and S-form shows the spheroidal structures (Reproduced with permission, Dunker et al. 1991a, b)

of conversion of water-soluble structure into an integral membrane oligomer required for membrane penetration. MBP is a major constituent of the multilamellar membranous myelin sheath surrounding the nerve axons. The water soluble form of myelin has also been shown to exist in intrinsic coil structure. However, its binding to lipids transforms the protein into the MG state (Polverini et al. 1999). Other examples, include retinal phosphodiesterase which also shows folding into a MG-state upon self-association of the intrinsically folded γ-subunit of the protein (Uversky et al. 2002). The coil structure of the fragment (77–262) of the glucocorticoid receptor (Baskakov et al. 1999) and N-terminal of HIV-1 integrase (Zheng et al. 1996) also undergoes structural rearrangements in the presence of trimethylamine N-oxide and zinc, respectively, leading to MG state formation.

4.6 Genetic Diseases

Folding intermediates can be introduced as a result of misfolding of translated polypeptides resulting from gene mutations. Such mutations eventually lead to the formation of mutant proteins which are not able to fold completely under normal physiological conditions (Bychkova and Ptitsyn 1995). Certain proteins (for example, beta-lactamase PC1 from *Staphylococcus aureus*, the N-terminal domain of lambda repressor) have been reported to form such intermediates due to the introduction of point mutations. In the case of genetic diseases such as cystic fibrosis, point mutations are often related to mislocation of proteins (Yang et al. 1993). The MG state is either involved at the intermediate stages of the unfolding of proteins or it could also act as the last stage leading to the closure of the unfolding process (Craig et al. 1985; Lim et al. 1992). Bychkova and Ptitsyn (1995) have also suggested that such mutations which cause genetic diseases by changing intracellular pathways of proteins could also act as the last step of protein folding. Hence, MG states of such proteins are either associated with chaperones or would aggregate, thus preventing them from interfering in the normal trafficking in the cell (Ptitsyn et al. 1995). The last action of MG could be helping in location and eventual degradation of mutated proteins. However, the protein conformational diseases might also occur due to misidentification or missignaling other than protein misfolding. The involvement of protein misfolding in such diseases could be beneficial in the design and development of drugs in forms of ligands or factors required for correct conformation of these proteins.

4.7 Conclusions

Cell signaling, human diseases, and intrinsic disorder are interrelated, perhaps having intriguing connections. It is now well-known that non-native conformations also play important biological roles in the functioning of cells along with the native

states. Membrane transport of proteins provides an exceptional instance of the mechanisms by which the cell engineers the process to achieve desired results by manipulating the structure of the protein. The MG state of proteins is believed to be part of the cell engineering which forms important intermediate conformations required for protein folding. The non-native conformations not only involve transmembrane translocation but also become a part of processes taking place within the cells. The MG state exists as a part of these non-native conformations. The MG conformations are required for various biological processes including translocation of proteins across the organelles or cell membranes, assisting molecular chaperones, binding to proteins to prevent them from aggregating until the quaternary structure is formed, ligand transfer, binding to multiple ligands, protein degradation at the acidic pH, channel or pore formation under low pH conditions, and for enzymatic activity. Proteins that form MGs rather than ordered structure under physiological conditions continue to be discovered. Another interesting role of MG state includes genetic diseases which occur due to mutations. Genetic diseases are mostly caused by misfolding of translated polypeptides that lead to increase in mutant proteins production resulting in various non-native conformations. The MG state definitely plays some role in such genetic diseases owing to their existence in the denatured state under physiological conditions. However, the diverse nature of non-native conformations can also not be ignored which involves conformations other than a MG. More native MG structures need to be unraveled to better understand the unfolding process of proteins. The growing database of bacterial as well as human genomes will help in the further understanding of native MG structures.

Various members of toxin family have demonstrated to acquire the functional state by transforming their native conformation to molten globule conformation. The vital role of the MG structure is significant in translocation of various toxins in view of their exquisite specificity of their mechanism of actions. The remarkable utility of unraveling of MG state in the steps of intoxication will help in designing of effective diagnostics and specific antidotes against their effects. More information about such states will also help in unraveling of unknown folding pathways as well as enzymatic activities. Careful investigation of MGs in toxins is necessary to unravel various unknown and the less clear phenomenon of protein chemistry. This family of proteins provides an integrated model to study various aspects of protein chemistry including factor affecting structural conformations of a protein, specificity, binding, translocation, longevity and various cellular functions.

References

Abelev GI (1993) Alpha-fetoprotein biology. Sov Sci Rev D Physicochem Biol 11:85–109

Allured VS, Collier RJ, Carroll SF et al (1986) Structure of exotoxin A of Pseudomonas aeruginosa at 3.0-Angstrom resolution. Proc Natl Acad Sci U S A 83:1320–1324

Arnon SS, Schechter R, Inglesby TV et al (2001) Botulinum toxin as a biological weapon, medical and public health management. JAMA 285:1059–1070

Ausio J, Dong F, van Holde KE (1989) Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "Tails" in the stabilization of the nucleosome. J Mol Biol 206:451–463
Bailey RW, Dunker AK, Brown CJ et al (2001) Clusterin, a binding protein with a molten globule-like region. Biochemistry 40:11828–11840
Bailey R, Griswold MD (1999) Clusterin in the male reproductive system, localization and possible function. Mol Cell Endocrinol 151:17–23
Bakás L, Chanturiya A, Herlax V et al (2006) Paradoxical lipid dependence of pores formed by the Escherichia coli alpha-hemolysin in planar phospholipid bilayer membranes. Biophys J 91:3748–3755
Bakás L, Ostolaza H, Vaz WL et al (1996) Reversible adsorption and nonreversible insertion of Escherichia coli alpha-hemolysin into lipid bilayers. Biophys J 71:1869–1876
Bakás L, Veiga M, Soloaga A et al (1998) Calcium-dependent conformation of E. coli alpha-haemolysin. Implications for the mechanism of membrane insertion and lysis. Biochim Biophys Acta 1368:225–234
Bantel H, Sinha B, Domschke W et al (2001) α-Toxin is a mediator of Staphylococcus aureus-induced cell death and activates caspases via the intrinsic death pathway independently of death receptor signaling. J Cell Biol 155:637–648
Baskakov IV, Kumar R, Srinivasan G et al (1999) Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. J Biol Chem 274:10693–10696
Bayer TS, Booth LN, Knudsen SM et al (2005) Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA 11:1848–1857
Bhakdi S, Tranum-Jensen J (1991) Alpha-toxin of Staphylococcus aureus. Microbiol Rev 55:733–751
Bose HS, Whittal RM, Baldwin MA et al (1999) The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. Proc Natl Acad Sci U S A 96:7250–7255
Bourdineaud JP, Boulanger P, Lazdunski C et al (1990a) In vivo properties of colicin A, channel activity is voltage dependent but translocation may be voltage independent. Proc Natl Acad Sci U S A 87:1037–1041
Bourdineaud JP, Fierobe HP, Lazdunski C et al (1990b) Involvement of OmpF during reception and translocation steps of colicin N entry. Mol Microbiol 4:1737–1743
Bova MP, Ding LL, Horwitz J et al (1997) Subunit exchange of alphaA-crystallin. J Biol Chem 272:29511–29517
Bumba L, Masin J, Fiser R et al (2010) Bordetella adenylate cyclase toxin mobilizes its b2 integrin receptor into lipid rafts to accomplish translocation across target cell membrane in two steps. PLoS Pathog 6:e1000901
Burge S, Teufel DP, Townsley FM et al (2009) Molecular basis of the interactions between the p73 N terminus and p300, effects on transactivation and modulation by phosphorylation. Proc Natl Acad Sci U S A 106:3142–3147
Bychkova VE, Berni R, Rossi GL et al (1992) Retinol-binding protein is in the molten globule state at low pH. Biochemistry 31:7566–7571
Bychkova VE, Dujsekina AE, Fantuzzi A et al (1998) Release of retinol and denaturation of its plasma carrier, retinol-binding protein. Fold Des 3:285–291
Bychkova VE, Dujsekina AE, Klenin SI et al (1996) Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface. Biochemistry 35:6058–6063
Bychkova VE, Pain RH, Ptitsyn OB (1988) The 'molten globule' state is involved in the translocation of proteins across membranes? FEBS Lett 238:231–234
Bychkova VE, Ptitsyn OB (1993) The molten globule in vitro and in vivo. Chemtracts Biochem Mol Biol 4:133–163
Bychkova VE, Ptitsyn OB (1995) Folding intermediates are involved in genetic diseases? FEBS Lett 359:6–8

Cabiaux V, Brasseur R, Wattiez R et al (1989) Secondary structure of diphtheria toxin and its fragments interacting with acidic liposomes studied by polarized infrared spectroscopy. J Biol Chem 264:4928–4938

Cai S, Kukreja R, Shoesmith S et al (2006) Botulinum neurotoxin light chain refolds at endosomal pH for its translocation. Protein J 25:455–462

Cai S, Singh BR (2001) Role of the disulfide cleavage induced molten globule state of type a botulinum neurotoxin in its endopeptidase activity. Biochemistry 50:15327–15333

Cai Z, Gorin A, Frederick R et al (1998) Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. Nat Struct Biol 5:203–212

Calnan BJ, Biancalana S, Frankel AD (1991) Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. Genes Dev 5:201–210

Casu F, Duggan BM, Hennig M (2013) The arginine-rich RNA-binding motif of HIV-1 Rev is intrinsically disordered and folds upon RRE binding. Biophys J 105:1004–1017

Cavalieri SJ, Bohach GA, Snyder IS (1984) Escherichia coli alpha-hemolysin, characteristics and probable role in pathogenicity. Microbiol Rev 48:326–343

Chahal SS, Matthews HR, Bradbury EM (1980) Acetylation of histone H4 and its role in chromatin structure and function. Nature 287:76–79

Chen E, Van Vranken V, Kliger DS (2008) The folding kinetics of the SDS-induced molten globule form of reduced cytochrome c. Biochemistry 47:5450–5459

Chenal A, Guijarro JI, Raynal B et al (2009) RTX calcium binding motifs are intrinsically disordered in the absence of calcium, implication for protein secretion. J Biol Chem 284:1781–1789

Chenal A, Karst JC, Sotomayor Pérez AC et al (2010) Calcium-induced folding and stabilization of the intrinsically disordered RTX domain of the CyaA toxin. Biophys J 99:3744–3753

Choe S, Bennett MJ, Fujii G et al (1992) The crystal structure of diphtheria toxin. Nature 357:216–222

Collier RJ, Young JA (2003) Anthrax toxin. Annu Rev Cell Dev Biol 19:45–70

Cortajarena A, Gon˜i FM, Ostolaza H (2001) Glycophorin as a receptor for Escherichia coli R-hemolysin in erythrocytes. J Biol Chem 276:12513–12519

Cotte JG (1992) Structural and functional relationships among the RTX toxin determinants of Gram negative bacteria. FEMS Microbiol Rev 88:137–161

Craig S, Hollecker M, Creighton TE et al (1985) Single amino acid mutations block a late step in the folding of β-lactamase from Staphylococcus aureus. J Mol Biol 185:681–687

Craig S, Schmeissner U, Wingfield P et al (1987) Conformation, stability, and folding of interleukin 1 beta. Biochemistry 26:3570–3576

de Jongh HH, Killian JA, de Kruijff B (1992) A water–lipid interface induces a highly dynamic folded state in apocytochrome c and cytochrome c, which may represent a common folding intermediate. Biochemistry 31:1636–1643

de Silva HV, Harmony JA, Stuart WD et al (1990) Apolipoprotein J, structure and tissue distribution. Biochemistry 29:5380–5389

Debaisieux S, Rayne F, Yezid H et al (2012) The ins and outs of HIV-1 Tat. Traffic 13:355–363

Demarest SJ, Deechongkit S, Dyson HJ et al (2004) Packing, specificity, and mutability at the binding interface between the p160 coactivator and CREB binding protein. Protein Sci 13:203–210

Demarest SJ, Martinez-Yamout M, Chung J et al (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature 415:549–553

Denton JB, Konishi Y, Scheraga HA (1982) Folding of ribonuclease A from a partially disordered conformation. Kinetic study under folding conditions. Biochemistry 21:5155–5163

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

Dolgikh DA, Abaturov LV, Bolotina IA et al (1985) Compact state of a protein molecule with pronounced small-scale mobility, bovine alpha-lactalbumin. Eur Biophys J 13:109–121

Dolgikh DA, Abaturov LV, Brazhnikov et al (1983) Acid form of carbonic anhydrase, “molten globule” with a secondary structure. Dokl Akad Nauk SSSR 272:1481–1484

Dolgikh DA, Gilmanshin RI, Brazhnikov EV et al (1981) Alpha-lactalbumin, compact state with fluctuating tertiary structure? FEBS Lett 136:311–315
Dolgikh DA, Kolomiets AP, Bolotina IA et al (1984) 'Molten-globule' state accumulates in carbonic anhydrase folding. FEBS Lett 165:88–92
D'Silva PR, Lala AK (1998) Unfolding of diphtheria toxin. Identification of hydrophobic sites exposed on lowering of pH by photolabeling. J Biol Chem 273:16216–16222
Dumont ME, Richards FM (1988) The pH-dependent conformational change of diphtheria toxin. J Biol Chem 263:2087–2097
Dunker AK, Ensign LD, Arnold GE et al (1991a) A model for fd phage penetration and assembly. FEBS Lett 292:271–274
Dunker AK, Ensign LD, Arnold GE et al (1991b) Proposed molten globule intermediate in fd phase penetration and assembly. FEBS Lett 292:275–278
Dunker AK, Lawson JD, Brown CJ (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59
Ebert MO, Bae SH, Dyson HJ et al (2008) NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR. Biochemistry 47:1299–1308
Eilers M, Endo T, Schatz G (1989) Adriamycin, a drug interacting with acidic phospholipids, blocks import of precursor proteins by isolated yeast mitochondria. J Biol Chem 264:2945–2950
Eliers M, Hwang S, Schatz G (1988) Unfolding and refolding of a purified precursor protein during import into isolated mitochondria. EMBO J 7:1139–1145
Eisenberg M, Gresalfi T, Riccio T et al (1979) Absorption of monovalent cations to bilayer membranes containing negative phospholipids. Biochemistry 18:5213–5223
Eliezer D, Jennings PA, Dyson HJ et al (1997) Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR. FEBS Lett 417:92–96
Endo T, Eilers M, Schatz G (1989) Binding of a tightly folded artificial mitochondrial precursor protein to the mitochondrial outer-membrane involves a lipid-mediated conformational change. J Biol Chem 264:2951–2956
Endo T, Schatz G (1988) Latent membrane perturbation activity of a mitochondrial precursor protein is exposed by unfolding. EMBO J 7:1153–1158
Farahbakhsh ZT, Wisnieski BJ (1989) The acid-triggered entry pathway of Pseudomonas exotoxin A. Biochemistry 28:580–585
Forestier C, Welch RA (1991) Identification of RTX toxin target cell specificity domains by use of hybrid genes. Infect Immun 59:4212–4220
Freskgård PO, Bergenhem N, Jonsson BH et al (1992) Isomerase and chaperone activity of prolyl isomerase in the folding of carbonic anhydrase. Science 258:466–468
Frey J (2011) The role of RTX toxins in host specificity of animal pathogenic Pasteurellaceae. Vet Microbiol 153:51–58
Fridd SL, Lakey JH (2002) Surface aspartate residues are essential for the stability of colicin A P-domain, a mechanism for the formation of an acidic molten-globule. Biochemistry 41:1579–1586
Gatewood JM, Schroth GP, Schmid CW et al (1990) Zinc-induced secondary structure transitions in human sperm protamines. J Biol Chem 265:20667–20672
Gill DM (1982) Bacterial toxins, A table of lethal amounts. Microbiol Rev 46:86–94
Gil'manshin RI, Dolgikh DA, Ptitsyn OB et al (1982) Protein globule without the unique three-dimensional structure, experimental data for alpha-lactalbumins and general model. Biofizika 27:1005–1016
Gil'manshin RI, Ptitsyn OB (1987) An early intermediate of refolding alpha-lactalbumin forms within 20 ms. FEBS Lett 223:327–329
González-Mañas JM, Lakey JH, Pattus F (1992) Brominated phospholipids as a tool for monitoring the membrane insertion of colicin A. Biochemistry 31:7294–7300

Griffith J, Manning M, Dunn K (1981) Filamentous bacteriophage contract into hollow spherical particles upon exposure to a chloroform–water interface. Cell 23:747–753
Griswold MD (1995) Interactions between germ cells and Sertoli cells in the testis. Biol Reprod 52:211–216
Grottesi A, Sette M, Palamara T (1998) The conformation of peptide thymosin alpha 1 in solution and in a membrane-like environment by circular dichroism and NMR spectroscopy. A possible model for its interaction with the lymphocyte membrane. Peptides 19:1731–1738
Gu W, Shi XL, Roeder RG (1997) Synergistic activation of transcription by CBP and p53. Nature 387:819–823
Hanski E, Farfel Z (1985) Bordetella pertussis invasive adenylate cyclase. Partial resolution and properties of its cellular penetration. J Biol Chem 260:5526–5532
Hartl FU, Neupert W (1990) Protein sorting to mitochondria, evolutionary conservations of folding and assembly. Science 247:930–938
Haslbeck M, Walke S, Stromer T et al (1999) Hsp26, a temperature-regulated chaperone. EMBO J 18:6744–6751
Henkens RW, Kitchell BB, Lottich SC et al (1982) Detection and characterization using circular dichroism and fluorescence spectroscopy of a stable intermediate conformation formed in the denaturation of bovine carbonic anhydrase with guanidinium chloride. Biochemistry 21:5918–5923
Herlax V, Bakas L (2007) Fatty acids covalently bound to alpha-hemolysin of Escherichia coli are involved in the molten globule conformation, implication of disordered regions in binding promiscuity. Biochemistry 46:5177–5184
Herlax V, Maté S, Rimoldi O et al (2009) Relevance of fatty acid covalently bound to Escherichia coli alpha-hemolysin and membrane microdomains in the oligomerization process. J Biol Chem 284:25199–25210
Humphreys DT, Carver JA, Easterbrook-Smith SB et al (1999) Clusterin has chaperone-like activity similar to that of small heat shock proteins. J Biol Chem 274:6875–6881
Ikeguchi M, Kuwajima K, Mitani M et al (1986) Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate, a comparative study of the folding reactions of alpha-lactalbumin and lysozyme. Biochemistry 25:6965–6972
Iram A, Naeem A (2012) Trifluoroethanol and acetonitrile induced formation of the molten globule states and aggregates of cellulase. Int J Biol Macromol 50:932–938
Janowiak BE, Fischer A, Collier RJ (2010) Effects of introducing a single charged residue into the phenylalanine clamp of multimeric anthraxprotective antigen. J Biol Chem 285:8130–8137
Jiang JX, Abrams FS, London E (1991) Folding changes in membrane-inserted diphtheria toxin that may play important roles in its translocation. Biochemistry 30:3857–3864
Jiang JX, London E (1990) Involvement of denaturation-like changes in Pseudomonas exotoxin a hydrophobicity and membrane penetration determined by characterization of pH and thermal transitions. Roles of two distinct conformationally altered states. J Biol Chem 265:8636–8641
Karn J, Dingwall C, Finch JT et al (1991) RNA binding by the tat and rev proteins of HIV-1. Biochimie 73:9–16
Kjaergaard M, Teilum K, Poulsen FM (2010) Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. PNAS 107:12535–12540
Koide S, Dyson HJ, Wright PE (1993) Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. Biochemistry 32:12299–12310
Krantz BA, Finkelstein A, Collier RJ (2006) Protein translocation through the anthrax toxin transmembrane pore is driven by a proton gradient. J Mol Biol 355:968–979
Krantz BA, Melnyk RA, Zhang S et al (2005) A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. Science 309:777–781
Krantz BA, Trivedi AD, Cunningham K et al (2004) Acid-induced unfolding of the amino-terminal domains of the lethal and edema factors of anthrax toxin. J Mol Biol 344:739–756
Kukreja RV, Singh BR (2005) Biologically active novel conformational state of botulinum, the most poisonous poison. J Biol Chem 280:39346–39352

Kumar R, Kukreja RV, Cai S et al (2014a) Differential role of molten globule and protein folding in distinguishing unique features of botulinum neurotoxin. Biochim Biophys Acta 1844:1145–1152
Kumar T, Sharma GS, Singh LR (2014b) Existence of molten globule state in homocysteine-induced protein covalent modifications. PLoS One 9:e113566
Kuwajim K (1996) The molten globule state of alpha lactalbumin. FASEB J 10:102–109
Kuwajima K, Hiraoka Y, Ikeguchi M et al (1985) Comparison of the transient folding intermediates in lysozyme and alpha-lactalbumin. Biochemistry 24:874–881
Lacy DB, Tepp W, Cohen AC et al (1998) Crystal structure of botulinum neurotoxin type A and implications for toxicity. Nat Struct Biol 5:898–902
Ladant D, Ullmann A (1999) Bordatella pertussis adenylate cyclase, a toxin with multiple talents. Trends Microbiol 7:172–176
Lakey JH, González-Mañas JM, van der Goot FG et al (1992) The membrane insertion of colicins. FEBS Lett 307:26–29
Lally ET, Hill RB, Kieba IR et al (1999) The interaction between RTX toxins and target cells. Trends Microbiol 7:356–361
Lally ET, Kieba IR, Sato A et al (1997) RTX toxins recognize a beta2 integrin on the surface of human target cells. J Biol Chem 272:30463–30469
Laminet AA, Ziegelhoffer T, Georgopoulos C et al (1990) The Escherichia coli heat shock proteins GroEL and GroES modulate the folding of the beta-lactamase precursor. EMBO J 9:2315–2319
Lecker SH, Driessen AJM, Wickner W (1990) ProOmpA contains secondary and tertiary structure prior to translocation and is shielded from aggregation by association with SecB protein. EMBO J 9:2309–2314
Li H, Chen JD (1998) The receptor-associated coactivator 3 activates transcription through CREB-binding protein recruitment and autoregulation. J Biol Chem 273:5948–5954
Li JD, Carroll J, Ellar DJ (1991) Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 Å resolution. Nature 353:815–821
Li HM, Theg SM, Bauerle CM et al (1990) Metal-ion-center assembly of ferredoxin and plastocyanin in isolated chloroplasts. Proc Natl Acad Sci U S A 87:6748–6752
Lim WA, Farrugio DC, Sauer RT (1992) Structural and energetic consequence of disruptive mutations in a protein core. Biochemistry 31:4324–4333
Lin RT, Genin P, Mamane Y et al (2000) Selective DNA binding and association with the CREB binding protein coactivator contribute to differential activation of alpha/beta interferon genes by interferon regulatory factors 3 and 7. Mol Cell Biol 20:6342–6353
Lin RT, Heylbroeck C, Pitha PM et al (1998) Virus-dependent phosphorylation of the IRF-3 transcription factor regulates nuclear translocation, transactivation potential, and proteasome-mediated degradation. Mol Cell Biol 18:2986–2996
Linhartová I, Bumba L, Mašín J et al (2010) RTX proteins, a highly diverse family secreted by a common mechanism. FEMS Microbiol Rev 34:1076–1112
Lohner K, Esser AF (1991) Thermal unfolding and aggregation of human complement protein C9, a differential scanning calorimetry study. Biochemistry 30:6620–6625
London E (1992a) Diphtheria toxin, membrane interaction and membrane translocation. Biochim Biophys Acta 1113:25–51
London E (1992b) How bacterial protein toxins enter cells; the role of partial unfolding in membrane translocation. Mol Microbiol 6:3277–3282
López G, Bañares-Hidalgo A, Estrada P (2011) Xylanase II from Trichoderma reesei QM 9414, conformational and catalytic stability to Chaotropes, Trifluoroethanol, and pH changes. J Ind Microbiol Biotechnol 38:113–125
Luger K, Mader AW, Richmond RK et al (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389:251–260
Luger K, Richmond TJ (1998) The histone tails of the nucleosome. Curr Opin Genet Dev 8:140–146
Manning M, Chryosogelos S, Griffith J (1981) Mechanism of coliphage M13 contraction, intermediate structures trapped at low temperatures. J Virol 40:912–991

Manning M, Chryosogelos S, Griffith J (1982) Insertion of bacteriophage M13 coat protein into membranes. Biophys J 37:28–30
Manning M, Griffith J (1985) Association of M13 I forms and spheroids with lipid vesicles. Arch Biochem Biophys 236:297–303
Marion J, Trovaslet M, Martinez N et al (2015) Pressure-induced molten globule state of human acetylcholinesterase, structural and dynamical changes monitored by neutron scattering. Phys Chem Chem Phys 17:3157–3163
Markus MA, Hinck AP, Huang SR et al (1997) High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. Nat Struct Biol 4:70–77
Martin J, Langer T, Boteva R et al (1991) Chaperonin-mediated protein folding at the surface of groEL through a 'molten globule'-like intermediate. Nature 352:36–42
Mattson MP, Shea TB (2003) Folate and homocysteine metabolism in neural plasticity and neurodegenerative disorders. Trends Neurosci 26:137–146
Merchant S, Bogorad L (1986) Rapid degradation of apoplastocyanin in Cu(II)-deficient cells of Chlamydomonas reinhardtii. J Biol Chem 261:15850–15853
Merrill AR, Cohen FS, Cramer WA (1990) On the nature of the structural change of the colicin E1 channel peptide necessary for its translocation-competent state. Biochemistry 29:5829–5836
Mishra P, Choudhary S, Hosur RV (2015) Ribosomal Protein P2 from apicomplexan parasite Toxoplasma gondii is intrinsically a molten globule. Biophys Chem 200–201:27–33
Mizejewski GJ (2004) Biological roles of alpha-fetoprotein during pregnancy and perinatal development. Exp Biol Med (Maywood) 229:439–463
Mizzen CA, Allis CD (1998) Linking histone acetylation to transcriptional regulation. Cell Mol Life Sci 54:6–20
Naeem A, Khan RH (2004) Characterization of molten globule state of cytochrome c at alkaline, native and acidic pH induced by butanol and SDS. Int J Biochem Cell Biol 36:2281–2292
Nassi S, Collier RJ, Finkelstein A (2002) PA63 channel of anthrax toxin, an extended beta-barrel. Biochemistry 41:1445–1450
Nishimura C, Dyson HJ, Wright PE (2008) The kinetic and equilibrium molten globule intermediates of apoleghemoglobin differ in structure. J Mol Biol 378:715–725
Obeid R, Herrmann W (2006) Mechanisms of homocysteine neurotoxicity in neurodegenerative diseases with special reference to dementia. FEBS Lett 580:2994–3005
Ohgushi M, Wada A (1983) 'Molten-globule state', a compact form of globular proteins with mobile side-chains. FEBS Lett 164:21–24
Ohgushi M, Wada A (1984) Liquid-like state of side chains at the intermediate stage of protein denaturation. Adv Biophys 18:75–90
Oliva R, Bazett-Jones DP, Locklear L et al (1990) Histone hyperacetylation can induce unfolding of the nucleosome core particle. Nucleic Acids Res 18:2739–2747
Ostolaza H, Bakás L, Goñi FM (1997) Balance of electrostatic and hydrophobic interactions in the lysis of model membranes by E. coli alpha-haemolysin. J Membr Biol 158:137–145
Ostolaza H, Soloaga A, Gon˜i FM (1995) The binding of divalent cations to Escherichia coli alpha-haemolysin. Eur J Biochem 228:39–44
Paoli P, Sbrana F, Tiribilli B et al (2010) Protein N-homocysteinylation induces the formation of toxic amyloid-like protofibrils. J Mol Biol 400:889–907
Papini E, Schiavo G, Tomasi M et al (1987) Lipid interaction of diphtheria toxin and mutants with altered fragment B. 2. Hydrophobic photolabelling and cell intoxication. Eur J Biochem 169:637–644
Parker MW, Pattus F, Tucker AD et al (1989) Structure of the membrane-pore-forming fragment of colicin A. Nature 337:93–96
Pathak BK, Mondal S, Ghosh AN et al (2014) The ribosome can prevent aggregation of partially folded protein intermediates, studies using the *Escherichia coli* ribosome. PLoS One 9:e96425
Polverini E, Fasano A, Zito F et al (1999) Conformation of bovine myelin basic protein purified with bound lipids. Eur Biophys J 28:351–355

Poon S, Easterbrook-Smith SB, Rybchyn MS et al (2000) Clusterin is an ATP-independent chaperone with very broad substrate specificity that stabilizes stressed proteins in a folding-competent state. Biochemistry 39:15953–15960

Poon S, Rybchyn MS, Easterbrook-Smith SB et al (2002a) Mildly acidic pH activates the extracellular molecular chaperone clusterin. J Biol Chem 277:39532–39540

Poon S, Treweek TM, Wilson MR et al (2002b) Clusterin is an extracellular chaperone that specifically interacts with slowly aggregating proteins on their off-folding pathway. FEBS Lett 513:259–266

Ptitsyn OB (1992) The molten globule state. In: Creighton TE (ed) Protein folding. Freeman, New York, pp 243–299

Ptitsyn OB (1995) Molten globule and protein folding. Adv Protein Chem 47:83–229

Ptitsyn OB, Bychkova VE, Uversky VN (1995) Kinetic and equilibrium folding intermediates. Philos Trans R Soc Lond B Biol Sci 348:35–41

Ptitsyn OB, Uversky VN (1994) The molten globule is a third thermodynamical state of protein molecules. FEBS Lett 341:15–18

Ptitsyn OB, Dolgikh DA, Gil'manshin RI et al (1983) Fluctuating state of the protein globule. Mol Biol (Mosk) 17:569–576

Ptitsyn OB, Pain RH, Semisotnov GV et al (1990) Evidence for a molten globule state as a general intermediate in protein folding. FEBS Lett 262:20–24

Qin BY, Liu C, Srinath H, Lam SS et al (2005) Crystal structure of IRF-3 in complex with CBP. Structure 13:1269–1277

Radford SE, Dobson CM, Evans PA (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. Nature 358:302–307

Rahaman H, Alam Khan MK, Hassan MI et al (2015) Heterogeneity of equilibrium molten globule state of cytochrome c induced by weak salt denaturants under physiological condition. PLoS One 10:e0120465

Rajaraman K, Raman B, Ramakrishna T et al (1998) The chaperone-like alpha-crystallin forms a complex only with the aggregation-prone molten globule state of alpha-lactalbumin. Biochem Biophys Res Commun 249:917–921

Rajaraman K, Raman B, Rao CM (1996) Molten-globule state of carbonic anhydrase binds to the chaperone-like alpha-crystallin. J Biol Chem 271:27595–27600

Raman B, Ramakrishna T, Rao CM (1997) Effect of the chaperone-like alpha-crystallin on the refolding of lysozyme and ribonuclease A. FEBS Lett 416:369–372

Ramsay G, Freire E (1990) Linked thermal and solute perturbation analysis of cooperative domain interactions in proteins. Structural stability of diphtheria toxin. Biochemistry 29:8677–8683

Ramsay G, Montgomery D, Berger D et al (1989) Energetics of diphtheria toxin membrane insertion and translocation, calorimetric characterization of the acid pH induced transition. Biochemistry 28:529–533

Randall LL, Hardy SJ (1989) Unity in function in the absence of consensus in sequence, role of leader peptides in export. Science 243:1156–1159

Ren J, Kachel K, Kim H et al (1999) Interaction of diphtheria toxin T domain with molten globule-like proteins and its implications for translocation. Science 284:955–957

Roberts LM, Dunker AK (1993) Structural changes accompanying chloroform-induced contraction of the filamentous phage fd. Biochemistry 32:10479–10488

Rosenberg ME, Girton R, Finkel D et al (2002) Apolipoprotein J/clusterin prevents a progressive glomerulopathy of aging. Mol Cell Biol 22:1893–1902

Sánchez-Magraner L, Cortajarena AL, García-Pacios M et al (2010) Interdomain Ca^{2+} effects in Escherichia coli alpha-haemolysin, Ca^{2+} binding to the C-terminal domain stabilizes both C- and N-terminal domains. Biochim Biophys Acta 1798:1225–1233

Sánchez-Magraner L, Viguera AR, García-Pacios M et al (2007) The calcium-binding C-terminal domain of Escherichia coli alpha-hemolysin is a major determinant in the surface-active properties of the protein. J Biol Chem 282:11827–11835

Satchell KJ (2007) MARTX, multifunctional autoprocessing repeats-in-toxin toxins. Infect Immun 75:5079–5084

Schiavo G, Matteoli M, Montecucco C (2000) Neurotoxins affecting neuroexocytosis. Physiol Rev 80:717–766
Schindel C, Zitzer A, Schulte B et al (2001) Interaction of Escherichia coli hemolysin with biological membranes. A study using cysteine scanning mutagenesis. Eur J Biochem 268:800–808
Schmidt MF, Robinson JP, DasGupta BR (1993) Direct visualization of botulinum neurotoxin-induced channels in phospholipid vesicles. Nature 364:827–830
Schreiber SL (1991) Chemistry and biology of the immunophilins and their immunosuppressive ligands. Science 251:283–287
Scoggin KES, Ulloa A, Nyborg JK (2001) The oncoprotein tax binds the SRC-1-interacting domain of CBP/p300 to mediate transcriptional activation. Mol Cell Biol 21:5520–5530
Semisotnov GV, Rodionova NA, Kutyshenko VP et al (1987) Sequential mechanism of refolding of carbonic anhydrase B. FEBS Lett 224:9–13
Sertoli E (1865) e lesistenza di particulari cellule remificate nei canalicoli seminiferi dell'testicolo umano. Morgangni 7:31–40
Shojania S, O'Neil JD (2006) HIV-1 Tat Is a natively unfolded protein. The solution conformation and dynamics of reduced HIV-1 Tat-(1–72) by NMR spectroscopy. J Biol Chem 281:8347–8356
Short EC, Kurtz HJ (1971) Properties of the hemolytic activities of Escherichia coli. Infect Immun 3:678–687
Singh BR (2000) Intimate details of the most poisonous poison. Nat Struct Biol 7:617–619
Skulachev VP (1988) Membrane bioenergetics. Springer, Berlin, Heidelberg, pp 19–25
Soloaga A, Ostolaza H, Gon˜i FM et al (1996) Purification of Escherichia coli pro-haemolysin, and a comparison with the properties of mature alpha-haemolysin. Eur J Biochem 238:418–422
Soloaga A, Veiga MP, García-Segura LM et al (1999) Insertion of Escherichia coli alpha-haemolysin in lipid bilayers as a non-transmembrane integral protein, prediction and experiment. Mol Microbiol 31:1013–1024
Sotomayor-Pérez A, Ladant D, Chenal A (2015) Disorder-to-order transition in the CyaA toxin RTX domain, implications for toxin secretion. Toxins (Basel) 7:1–20
Stroylova YY, Chobert JM, Muronetz VI et al (2012) N-homocysteinylation of ovine prion protein induces amyloid-like transformation. Arch Biochem Biophys 526:29–37
Su L, Radek JT, Labeots LA et al (1997) An RNA enhancer in a phage transcriptional antitermination complex functions as a structural switch. Genes Dev 11:2214–2226
Surendran R, Herman P, Ching LJ (2004) HIV Rev self-assembly is linked to a molten-globule to compact structural transition. Biophys Chem 108:101–119
Thoren KL, Krantz BA (2011) The unfolding story of anthrax toxin translocation. Mol Microbiol 80:588–595
Thoren KL, Worden EJ, Yassif JM (2009) Lethal factor unfolding is the most force-dependent step of anthrax toxin translocation. Proc Natl Acad Sci U S A 106:21555–21560
Tomasi TB (1977) Structure and function of alpha-fetoprotein. Annu Rev Med 28:453–465
Trougakos IP, Gonos ES (2006) Regulation of clusterin/apolipoprotein J, a functional homologue to the small heat shock proteins, by oxidative stress in ageing and age-related diseases. Free Radic Res 40:1324–1334
Tsou CL (1995) Inactivation precedes overall molecular conformation changes during enzyme denaturation. Biochim Biophys Acta 1253:151–162
Tsuruta JK, Wong K, Fritz IB et al (1990) Structural analysis of sulphated glycoprotein 2 from amino acid sequence. Relationship to clusterin and serum protein 40, 40. Biochem J 268:571–578
Undas A, Perla J, Lacinski M et al (2004) Autoantibodies against N-homocysteinylated proteins in humans, implications for atherosclerosis. Stroke 35:1299–1304
Uversky VN (2003) A rigidifying union, the role of ligands in protein structure and stability. In: Pandalai SG (ed) Recent research developments in biophysics and biochemistry, vol 3. Transworld Research Network, Kerala, pp 711–745
Uversky VN, Dunker AK (2010) Understanding protein non-folding. Biochim Biophys Acta 1804:1231–1264

Uversky VN, Gillespie JR, Millett IS et al (2000b) Zn^{2+}-mediated structure formation and compaction of the "natively unfolded" human prothymosin. Biochem Biophys Res Comunun 267:663–668
Uversky VN, Kirkitadze MD, Narizhneva NV et al (1995) Structural properties of alphafeto protein from human cord serum, the protein molecule at low pH possesses all the properties of the molten globule. FEBS Lett 364:165–167
Uversky VN, Kutyshenko VP, Protasova NY et al (1996) Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. Protein Sci 5:1844–1851
Uversky VN, Li J, Fink AL (2001) Metal-triggered structural transformations, aggregation and fibril formation of human-synuclein. A possible molecular link between Parkinson's disease and heavy metal exposure. J Biol Chem 276:44284–44296
Uversky VN, Narizhneva NV (1998) Effect of natural ligands on the structural properties and conformational stability of proteins. Biochemistry (Mosc) 63:420–433
Uversky VN, Narizhneva NV, Ivanova TV et al (1997) Ligand-free form of human alphafetoprotein, evidence for the molten globule state. FEBS Lett 410:280–284
Uversky VN, Permyakov SE, Zagranichny VE et al (2002) Effect of zinc and temperature on the conformation of the γ subunit of retinal phosphodiesterase. A natively unfolded protein. J Proteome Res 1:149–159
Uversky VN, Semisotnov GV, Pain RH et al (1992) 'All-or-none' mechanism of the molten globule unfolding. FEBS Lett 314:89–92
Vamvaca K, Vögeli B, Kast P et al (2004) An enzymatic molten globule, efficient coupling of folding and catalysis. Proc Natl Acad Sci U S A 101:12860–12864
van der Goot FG, Gonzalez-Manas JM, Lakey JH et al (1991) A 'molten-globule' membrane-insertion intermediate of the pore-forming domain of colicin A. Nature 354:408–410
van der Goot FG, Lakey JH, Pattus F (1992) The molten globule intermediate for protein insertion or translocation through membranes. Trends Cell Biol 2:343–348
van der Vies SM, Viitanen PV, Gatenby AA et al (1992) Conformational states of ribulosebisphosphate carboxylase their interaction with chaperonin 60. Biochemistry 31:3635–3644
Vécsey-Semjén B, Möllby R, van der Goot FG (1996) Partial C-terminal unfolding is required for channel formation by staphylococcal a-toxin. J Biol Chem 271:8655–8660
Wang C, Zhang Q, Cheng Y et al (2010) Refolding of denatured/reduced lysozyme at high concentrations by artificial molecular chaperone-ion exchange chromatography. Biotechnol Prog 26:1073–1079
Waterborg JH, Matthews HR (1983) Intranuclear localization of histone acetylation in Physarum polycephalum and the structure of functionally active chromatin. Cell Biophys 5:265–279
Waters L, Yue B, Veverka V et al (2006) Structural diversity in p160/CREB-binding protein coactivator complexes. J Biol Chem 281:14787–14795
Weiss AA, Hewlett EL (1986) Virulence factors of Bordetella pertussis. Annu Rev Microbiol 40:661–686
Weissman JS, Hohl CM, Kovalenko O et al (1995) Mechanism of GroEL action, productive release of polypeptide from a sequestered position under GroES. Cell 83:577–587
Welch RA (1991) Pore-forming cytolysins of Gram-negative bacteria. Mol Microbiol 5:521–528
Welch RA (2001) RTX toxin structure and function, a story of numerous anomalies and few analogies in toxin biology. Curr Top Microbiol Immunol 257:85–111
Wiles TJ, Mulvey MA (2013) The RTX pore-forming toxin α-hemolysin of uropathogenic Escherichia coli. Future Microbiol 8:73–84
Wong KP, Tanford C (1973) Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions. J Biol Chem 248:8518–8523
Yang J, Dunker AK, Powers JR et al (2001) Beta-lactoglobulin molten globule induced by high pressure. J Agric Food Chem 49:3236–3243
Yang Y, Janich S, Cohn JA et al (1993) The common variant of cystic fibrosis transmembrane conductance regulator is recognized by hsp70 and degraded in a pre-Golgi nonlysosomal compartment. Proc Natl Acad Sci U S A 90:9480–9484

Yonath A, Franceschi F (1997) New RNA recognition features revealed in ancient ribosomal proteins. Nat Struct Biol 4:3–5

Zhao JM, London E (1986) Similarity of the conformation of diphtheria toxin at high temperature to that in the membrane-penetrating low-pH state. Proc Natl Acad Sci U S A 83:2002–2006

Zhao JM, London E (1988) Conformation and model membrane interactions of diphtheria toxin fragment A. J Biol Chem 263:15369–15377

Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci U S A 93:13659–13664

Zurdo J, Sanz JM, Gonzalez C et al (1997) The exchangeable yeast ribosomal acidic protein YP2beta shows characteristics of a partly folded state under physiological conditions. Biochemistry 36:9625–9635

Chapter 5
Evolution of Toxin

Raj Kumar and Bal Ram Singh

Keywords Evolution • Lateral gene transfer • Three finger toxins • Inhibitor cystein knots • Gene duplication • Venoms • Toxins

5.1 Introduction

Biological evolution is generally considered as the basis of diversity of life on Earth, and is expressed through genetic variation caused by random mutations. It is notable that the gene products, the proteins, are the functional biomolecules which are ultimately responsible for the changes in phenotypes responsible for natural selection under changed environment. However, physico-chemical and molecular bases of mutations and mechanism involved in linking molecular changes to the slow evolution process, remain unknown. One of the ways to explore the forces of evolutionary pressure and mechanisms involved in the molecular behavior to identify a functional group of molecules from different sources for comparative analysis. We have therefore identified a group of molecules, toxins, which are certainly a special group of highly effective biological molecules which are generally involved in the survival of the organisms possessing such elements in their system.

A substance, which can be harmful to plant/animals/humans at low concentration, can be termed as a toxin. The toxins can be small molecules, pesticides or proteins, which can cause disease through interaction with biological macromolecules. Their molecular weight can vary between 100 and 100,000 or more. Unlike chemical toxins, biological toxins are more diverse and toxic. Biological toxins are produced mainly by animals or bacteria, although plants and fungi are also known to produce toxins. The common name for animal toxins is venoms. Whereas for microbial toxins, produced by bacteria or fungi, are termed as exotoxin or endotoxin. For the understanding evolution of toxins, one needs to understand the evolution of these two toxins producing kingdoms.

A venom can be simply defined as a secretion produced by a gland in animal and delivered into the target animal through the infliction of wounds. The delivered material contains molecules that disrupt normal physiological processes in the targeted

R. Kumar, B.R. Singh, *Protein Toxins in Modeling Biochemistry*, SpringerBriefs in Biochemistry and Molecular Biology, DOI 10.1007/978-3-319-43540-4_5

Table 5.1 Comparison of exotoxins and endotoxins

Exotoxins	Endotoxins
Excreted by organisms, living cell	Integral part of cell wall
Found in both gram positive and gram negative bacteria	Found mostly in gram negative bacteria
It is a polypeptide	It is lipopolysaccharide complex
Highly antigenic	Weakly immunogenic
Highly toxic	Moderately toxic
Usually bind to specific receptors	No specific receptors

animal. This process is used by the venomous animal to facilitate feeding or for defense. Most animal venoms are a highly complex cocktail of bioactive compounds, containing proteins, peptides, salts, organic compounds, amino acids and neurotransmitters (Fry et al. 2009, 2012; Olivera 2002; Escoubus and King 2009). The proteinaceous components are generally most abundant in the delivered material.

On the other hand, microbial toxins are one of the most potent molecules known to us. They directly damage the host tissue and the immune system. They are generally classified as endo or exotoxins. Exotoxins are generally secreted by bacteria, whereas endotoxins remain part of the host and released after lysis of host cells (Table 5.1).

In this chapter, we will describe these two groups of toxins separately. Both groups offer interesting insights into pharmacology, drug discovery, immunology and structural biology. First, we will discuss the evolutionary characteristics of venoms and bacterial toxins. In the second part, we will discuss the mechanism of evolution of animal and bacterial toxins.

5.2 Animal Venoms and Bacterial Toxins

One of the most interesting aspects of venom science is that it offers an unparallel model for investigating the relationship between predator and prey, protein neofunctionalization, and influence of natural selection in molecular evolution. Venoms serve multiple functions in the animal kingdom. They act as a defense in some (such as lizards, fish, insects, etc.), foraging adaptation among trophically venomous taxa (mammals, lizards, snakes, spiders, etc.), and potentially for intraspecies conflicts (platypus). Animal venoms are rich sources of peptides and proteins with various physiological effects. Generally, venoms consists of numerous peptides, proteins and chemicals. This diversity of function makes venom an important candidate for evolutionary innovations (Table 5.2), and is one of the main reasons why they are found across the animal kingdoms (Fig. 5.1). To spread throughout the animal kingdom venoms utilize several innovative structures, such as the beak, fangs, teeth, spins, sprays, harpoons, etc. (Fry et al. 2009; Smith and Wheeler 2006).

Table 5.2 Animal toxins with their source and activity

Toxin activity	Toxin	Animal source
Vasodilation by kinin releasing activity	Blarinatoxin	Blarina
	Kallikrein	
	Peptidase S1 toxins	Toxicoferan reptiles
Vasodilation by kinin-like activity	Tachykinin-like peptides	Cephalopods and Hymenopteran insects
Inhibition of epinephrine-induced platelet aggregation	Type III phospholipase A_2	Auguimorph lizards
	Type IB phospholipase A_2	Snakes
Inhibition of ADP-induced platelet aggregation	Snake venom metalloproteinases	Snakes
	Type IB phospholipase A_2	
	Type IIA phospholipase A_2	
	Apyrase	Hymenopteran insects
Inhibition of collagen-induced platelet aggregation	C-type lectins	Snakes
	Snake venom metalloproteinases	
Platelet aggregation inhibition by GPIIb/IIIa receptor antagonism	Disintegrins	Snakes
	Mambin/dendroaspin	
Prothrombin activation	Factor V	Snakes
	Factor X	
	Snake venom metalloproteinases	
Thrombin activation	C-type lectins	Snakes
	Textilinins	
	Type IIA phospholipase A_2	
Fibro(geno)lytic activity	Snake venom metalloproteinases	Snakes
	Peptidase S1 toxins	Toxicoferan reptiles
Neurotoxin activity		
Sodium channel blockers	μ-Conotoxins	Cone snails
	μ-O-conotoxins	
	Cn-11	Scorpions
	Hainatoxin-I	Spiders
	Protoxin-II	
Sodium channel activators (site-4-toxins)	β-Toxins	Scorpions
	δ-Palutoxins	Spiders
	μ-Agatoxins	
Sodium channel prolongers	δ-Conotoxins	Cone snails
	Uncharacterized toxins	Irukandji jellyfish
	α-Toxins	Scorpions
	Sea anemone sodium channel inhibitory toxins	Sea anemones
	δ-Atracotoxins	Spiders
Potassium channel blockers	k-Conotoxins	Cone snails
	Apamin	Hymenopteran insects
	Short scorpion toxins	Scorpions
	Cnidaria Kunitz-type proteinase inhibitors	Sea anemones
	Sea anemone type 3	
	Dendrotoxins	Snakes
	k-Atracotoxins	Spiders
	CRISP toxins	Toxicoferan reptiles

(continued)

Table 5.2 (continued)

Toxin activity	Toxin	Animal source
Calcium channel blockers	ω-Conotoxins	Cone snails
	Calcicludine	Snakes
	Claciseptine/FS2	
	ω-Neurotoxins	Spiders
	CRISP toxins	Toxicoferan reptiles
Nicotinic receptor antagonists	α-Conotoxins	Cone snails
	α-Neurotoxins	Snakes
Muscarinic receptor antagonists	Uncharacterized toxins	Scorpions
	Phospholipase A_2 toxins	Snakes
	Type-A muscarinic toxins	
	Type-B muscarinic toxins	

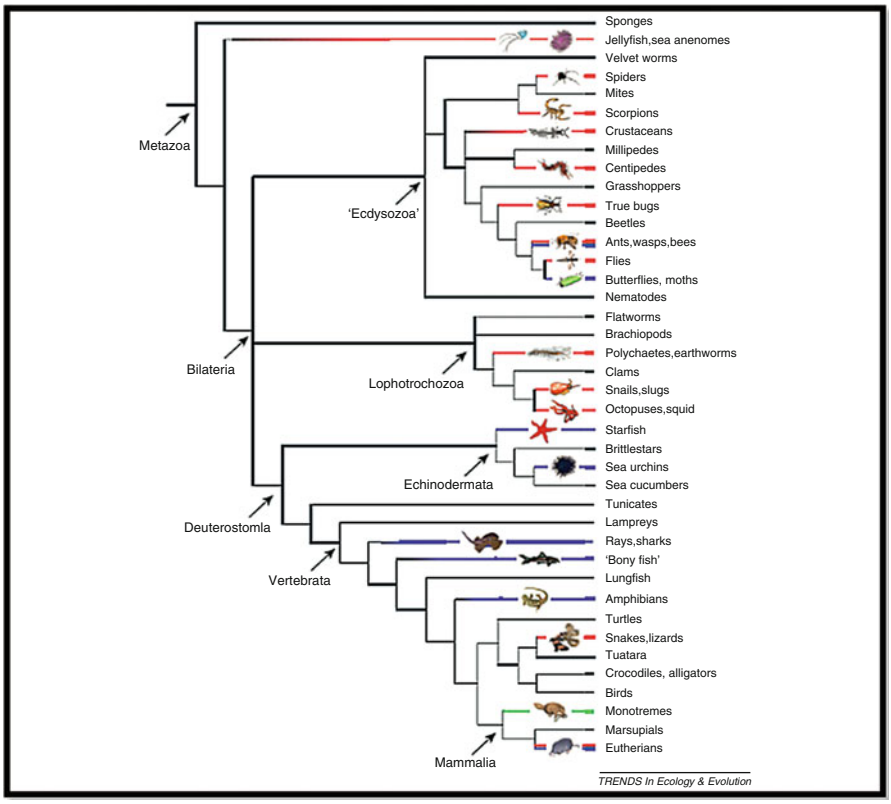

Fig. 5.1 Schematic tree of venomous life in animal kingdom (Reproduced with permission, Casewell et al. 2013)

The composition of venoms depends on their function and target. For example, defensive venoms are highly conserved whereas predatory venoms are more complex and variable in composition and action. On the other hand, bacteria harm the host by releasing toxins—proteins that are exceptionally effective poisons. Toxins are the most important virulence factor of bacterial harmful weaponry. Always targeting essential molecules, toxins typically go after molecules that are either scarce or whose role is to send important signals. This may be one of the reasons why bacterial toxins are most potent molecules on Earth. In both cases, only a small number of toxin molecules are required to cause the damage. In contrast, some toxins appear to deviate from these strategies by targeting highly abundant proteins. Most bacterial toxins have an A-B structure, B being a polypeptide which binds to the receptor and A being an enzyme. Many toxins are activated, either when produced by the bacterium or when bound to the membrane receptor, by proteases (nicking). Most toxins act on the target cells to which they bind (Botulinum Toxin), and to a lesser degree ascend axons and affect more distant structures (Tetanus Toxin). Several toxic effects caused by bacteria have been described. A few toxins have been identified, characterized, and their mode of action determined at the molecular level.

5.3 The Role of Natural Selection/Adaptive Evolution

Diversity in animals producing venoms as depicted in Fig. 5.1, and consideration of venom function, indicates towards natural selection. Biodiversity in animal and plants results in diversity in venoms. Although strong evidence is not present to suggest the venom composition as adaptive, the composition is correlated to diet (for example Malayan pit viper; Daltry et al. 1996). Also, in the natural environment, the amount of venom injected into prey is often greater than 100 times the lethal dose required (Mebs 2001). So it is difficult to establish the relationship between a component of a given venom and predator-prey co-evolution, due to varying composition of venoms.

Generally, gene duplication and functional divergence are responsible for the formation of functionally distinct genes. Interestingly these genes are evolutionarily related which is a fundamental process of adaptive evolution (Hughes 1994). Gene duplication is of significant biological importance in adaptive evolution where it can (i) provide a direct solution to the selection problem, (ii) facilitate further adaptation, and (iii) create new gene functions. Only the redundancy created by gene duplication allows these genes to escape the negative selection pressure and finally evolve into a new function. Once a duplicate gene has adapted to its specific function, purifying selection predominates. Purifying selection allows the number of synonymous substitutions (dS) to catch up to and eventually exceed the number of non-synonymous substitutions (dN). But in a limited period of time diversifying selection can be detected and in this time period, dN is larger than dS. The ratio between these rates ($\omega = dN/dS$) is the most common way to measure selection in

protein-coding genes. Comparison of the rates of synonymous and non-synonymous substitutions in duplicated genes can provide information on the relative role of natural selection and genetic drift in the functional diversification of genes. Values of $\omega = 1$, $\omega < 1$ and $\omega > 1$ indicate neutral selection, purifying and positive selection. An excess of d*N* over d*S* is a molecular indicator of adaptive evolution (Hughes and Nei 1998; Hughes 2000). When there is a strong purifying selection pressure (indicated by $\omega \ll 1$) to act upon a specific sequence region, there is a high sequence homology maintained by low ω value in that sequence region. There is always a negative correlation between ω value and sequence homology/similarity most of the time during the evolution. Alternatively, we can state that the ω value has the positive correlation with sequence diversity. In animal toxins, very often d*N* is larger than d*S*, indicative of positive Darwinian selection, such as PLA_2 (Ohno et al. 1998), PLA_2 inhibitor genes (Nobuhisa et al. 1997), three—finger toxin (Gong et al. 2000), serine proteinases (Deshimaru et al. 1996), and conotoxins (Duda and Palumbi 1999). ω value can be misleading in following cases (a) when recombination rates are high, (b) when bacteria/species have nucleotide composition biases (generally AT-rich), and (c) small effective population size. These cases lead to relaxed constraints on protein sequences, increase deleterious mutations resulting in inflated ω value (Toft and Andersson 2010). Adaptive evolution is not concentrated only on active sites, sometimes it operates to subdomains also (Nobuhisa et al. 1997).

Another example of adaptive evolution is botulinum toxin (BoNT). Calculations of the ω value of different serotypes suggest that there is no accelerated evolution rate in either among the different serotypes of BoNT or within the same serotype of BoNT. Interestingly, the scenario is quite different for the subunits of BoNT (light and heavy chains). BoNT/E light chain and TeNT heavy chain apparently have high ω value ($\omega > 1$) which means a strong positive selection pressure is acting on those regions. It will lead to a conclusion that BoNT/E light chain and TeNT heavy chain region are still evolving and have a faster evolution rate compared to the other regions or domains of this protein. Therefore, it is possible that the next novel serotype of endopeptidase (perhaps in the recognition of a new substrate) will evolve from BoNT/E and the next new receptor binding capability will evolve from TeNT (recognition of a new receptor). This is a great prediction of adaptability of toxin (Chang 2011).

5.3.1 Adaptive Evolution of Animal Toxin Structure

In general, three-dimensional or tertiary structures are more conserved than primary structure indicating that the fold is more important in the interactions. This also allows more variations in the primary structure during evolution. Compared to other proteins, in most animal toxins larger proportion of amino acids, particularly cysteines, are responsible for maintaining overall structures (in most proteins 3–4 % residues are responsible for maintaining the overall structure; Rost 1997).

In animal toxins, the existence of structural plasticity allows them to evolve as highly specific toxins for their respective targets. Specific binding of snake venoms

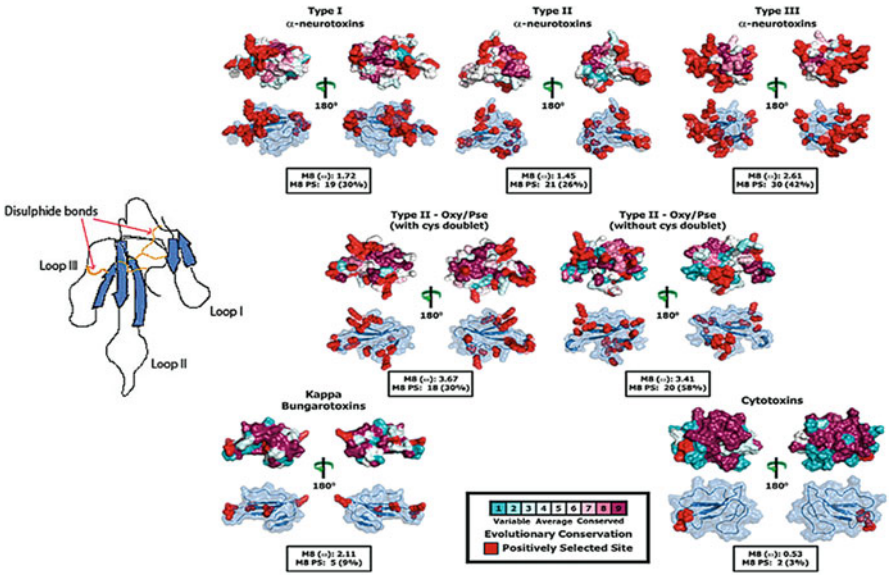

Fig. 5.2 Molecular evolution of derived 3FTx. Three dimensional homology model of various three-finger toxins. Three-dimensional homology models of various three-finger toxins (*top*: surface-fill; *bottom*: wireframe), depicting the locations of positively selected sites are presented. Site-model 8 computed omega and the total number of positively selected sites detected by its Bayes Empirical Bayes (BEB) approach ($PP \geq 0.95$) are also indicated. Disulphide bonds and arrangements of loops are shown (model picture on the *left*) (Reproduced with permission, Sunagar et al. 2013)

PLA_2 to its acceptors is due to the presence of pharmacological sites located on the surface which is other than the active sites (Kini 1997). These surface modifications have a significant role in the evolution of new pharmacological effects (Kini and Chan 1999). In all three-fingered toxins the core region consists of four disulfide bridges, the β-strands, and the c-terminal loops (Fig. 5.2). The overall structure is the same. Binding regions of different three-fingered toxins involve loop I and loop II (Fig. 5.2). Although they have similar fold but their different biological functions are achieved by the different arrangements of exposed loops (Menez 1998; Ohno et al. 1998). Most of the conformational differences among three-fingered toxins are due to the conformational differences in the loop regions. Comparison of several 3D structures revealed that the orientation of the loops can vary considerably, suggesting that these loops determine their biological specificity (Fig. 5.2).

Inhibitor cysteine knot (ICK) structural motif is a disulfide cross-linked polypeptide which has a different fold. This structure is characterized by three β-sheets stabilized by a cysteine knot (Figs. 5.2 and 5.3f). It is interesting that ICK fold is present in several toxins from organisms as diverse as spiders, snails, plants, and fungi. The structure–function studies of ICK motifs also demonstrated that most important regions for toxicity lie in the loop region (Nortan and Pallaghy 1998). Similarly, all scorpion toxins possess the same α/β "scorpion fold". They also have structural plasticity. Resistance to change in fold due to sequence mutation and

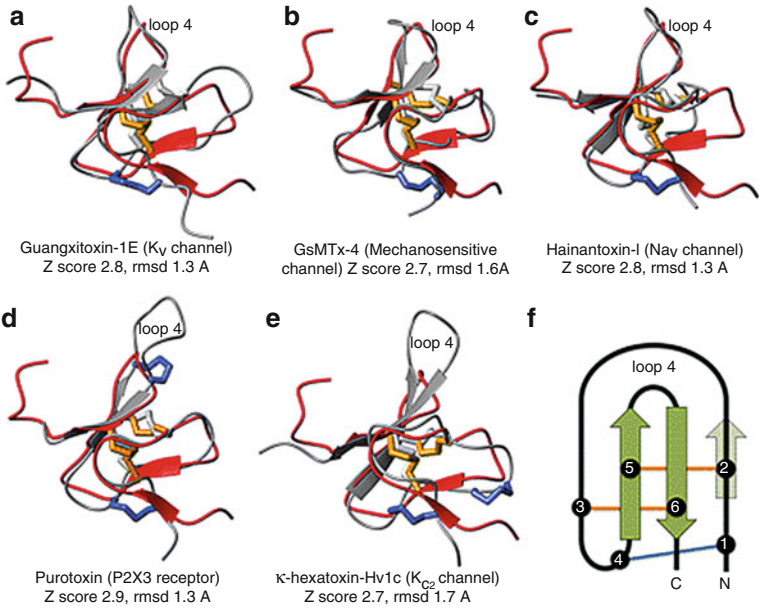

Fig. 5.3 Three-dimensional structure of various spider toxins. Overlay of U1-LITX-Lw1a and structurally homologous ICK toxins. The NMR structure of U1-LITX-Lw1a (*red*) is overlaid on the structures of different structurally homologous spider ICK toxins (*gray*), including (**a**) guangxitoxin (PDB ID code 2WH9), (**b**) GsMTX-4 (PDB ID code 1TYK), (**c**) hainantoxin—I (PDB ID code 1NIX), (**d**) purotoxin (PDB ID code 2KGU), and (**e**) k—hexatoxin—Hv1c (PDB ID code 1DL0). The two central disulfides of the ICK toxins, shown as *gray tubes*, overlap with the two disulphides of U1-LITX-Lw1a (shown as *gold tubes*). Additional disulfide bonds in the ICK toxins are shown as *blue tubes*. The molecular target of each spider toxin is indicated. (**f**) Graphical representation of the DDH and ICK motif shown in *orange* and *green*, respectively. The third disulphide necessary for the formation of the ICK motif is shown in *blue* (Reproduced with permission, Smith et al. 2011; Copyright (2008) National Academy of Sciences, USA)

multiple substitutions are common characteristics of animal toxins. These characteristics allow them to conserve their toxicity in different environments. It is not difficult to hypothesize that the structural plasticity is one of the characteristics needed for evolution.

5.4 Adaptive Evolution in Bacterial Toxins

Bacterial adaptation to new environmental conditions mainly involves two different mechanisms: alteration of gene regulation without any inheritable genetic change or selection of novel adaptive phenotypes conferred by stable mutations. In bacteria, gene duplication and amplification are the most frequent mutation type (Sandegren and Andersson 2009).

Another mechanism of bacterial adaptation is “chemotaxis”. This involves the ability to move toward nutrient (positive chemotaxis) and away from noxious com-

pounds (negative chemotaxis). Bacteria also have feed-back mechanisms, such as catabolite repression of transcription, in which an external signal is transduced through histidine protein kinases (Stock et al. 1990). These mechanisms help bacteria respond to changing the external environment and help with adaptation. Adaptive potential of pathogenic bacteria may depend on the host immune system. The pathogenic personality of bacteria, intracellular and intercellular locations, and antagonistic responses give shape to co-evolutionary processes. Because of a large population and short generation time, bacteria have evolutionary flexibility. So in changing environment, bacteria may respond through phenotypic variations or increasing the rate of their propensity.

Bacterial evolution is often considered as symbiotic evolution. They follow a co-evolutionary process with their hosts. There are two general models to explain the co-evolutionary process: (a) trench warfare model (Bergelson et al. 2001), (b) red queen model (Clay and Kover 1996). Although trench warfare model advocates polymorphism as the source of genetic diversity, but in principle both these models suggest that back and forth evolution between the bacterial protein and the targeted eukaryotic molecules. Positive selection on favorable alleles yields genetic variation over time, but at the same time reduces polymorphisms in the populations. One way to identify adaptation process at the molecular level is through estimation of ω (dN/dS) (discussed in Sect. 5.3). Another way involves functional divergence in some lineages in a particular protein region. Functional divergence is a process in which genes shift their function from and ancestral genes, after gene duplication. This leads to subfunctionalization, or nonfunctionalization through deleterious mutations. This process mostly uses a horizontal or lateral gene transfer (see further discussion later).

5.5 Mechanisms of Venom Evolution

There could be two ways an animal can acquire toxic substance: intrinsic or extrinsic. Animals carrying or showing toxin activity either use their own genetic material or that of other organisms for the production of these compounds. Gene expression leads to direct translation to a peptide or proteins which are toxic to targets. In a second way, animal produces some metabolites which will be transferred into a target, triggering numerous chemical reactions leads to secondary metabolites exhibiting allelochemical activity, such as toxic activity. Another way of acquiring toxins is through using toxic metabolites for own purpose or by having a symbiotic relationship, for example, microorganisms which produce toxins eventually used by the host. In this way, the animal need to develop resistance to the toxin's action by involving changes in ion channels, receptor, potential targets of toxin, and/or to develop a compartment where acquired toxin can be stored without risk of intoxication. Also, the organism needs to rely on a particular food source for proper metabolism in the presence of toxic material or provide appropriate conditions for accommodating toxin producing symbionts.

Diversity in toxin could be due to the gene pool which encodes a toxic protein and peptides. Point mutations, gene duplications, recombination, and post-

translational modification of genes lead to a wide variety of proteins and peptides. This diversity coupled with a limited change in structural frameworks is demonstrated in several venomous animal toxins (Figs. 5.2 and 5.3).

The complexity of snake venoms is associated with high variability of their components: enzymes and toxins. For example phospholipase A_2 changes their original function, digestion, and turned into highly lethal toxins like β-bungarotoxin, taipoxin or notexin (Kini and Chan 1999; Harris 1998). The venoms of the Australian brown snake are known to contain numerous isoenzymes of phospholipase A_2 (Takasaki et al. 1990a, b). Individual variability was demonstrated and each venom has its own sets of isoenzymes. Variability in venom composition occurs at all levels: interfamilial, intergeneric, intra- and interspecies (Chippaux et al. 1991). Gene duplication, accelerated evolution, gene conversion, and positive Darwinian selection contribute to the molecular evolution of PLA_2 enzymes. Natural substitutions occur primarily on the surface of PLA_2 to maintain similar folds but changing specificity. By altering the molecular surface, nature has developed a strategy to keep the same fold with different functions (Kini and Chan 1999; Table 5.2 and Figs. 5.2 and 5.3).

Venom system provides a unique opportunity to investigate the interrelations between natural selection and the genetic and molecular processes responsible for variability in species. Although gene duplication was suggested for toxin evolution and variability, it is not a prerequisite for the toxin recruitment. Similar to phospholipase A_2, several venom genes have been found to be homologous to gene loci that are physiologically expressed in non-venomous genes. Alternative splicing has also been proposed as a mechanism capable of generating non-venomous and venomous forms. There is a paradoxical phenomenon that venomous animals may have in common: the non-coding part of the toxin gene, the introns, represents highly conserved elements, whereas the coding parts, the exons exhibit great variability. The conservation of the intron sequences enables homologous recombination with other genes during meiosis, which contributes to even greater expansion. Another method for gene recruitment for toxin or venom gene is through alteration in the structure of domains. In this process, both single multidomain and multiple single domain products will be formed (Fig. 5.4).

In particular, positive selection appears to be a general phenomenon among venomous taxa (Duda and Palumbi 1999; Fry et al. 2003; Binford et al. 2009; Weinberger et al. 2011). The A-superfamily of conotoxin genes isolated from venomous contains some of the most rapidly evolving protein-coding genes identified in metazoans to date (Chang and Duda 2012).

Mainly three theories are proposed for diversity in multiple gene families: (a) Divergent evolution, (b) Concerted evolution, (c) Birth and Death model (Nei et al. 1997; Nei and Rooney 2005). Divergent evolution is the accumulation of differences between groups which can lead to the formation of new species. The mechanism of this model of evolution involves diffusion of the same species to a different and isolated environment, which blocks the gene flow among the distinct populations allowing the differentiated fixation of characteristics through genetic drift and natural selection. Divergent evolution can be applied to genes and proteins. This is possible with both orthologous genes (resulting from a speciation event) and paralogous

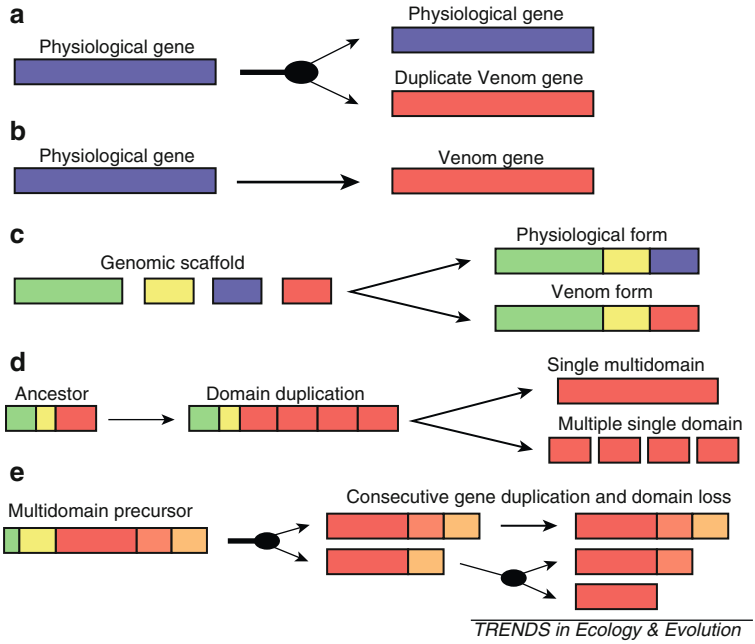

Fig. 5.4 Schematics showing the mechanism that underlies toxin evolution. (**a**) Gene duplication of a physiological gene and subsequent evolution of a duplicate gene into a venom toxin, (**b**) modification of a physiological gene into venom genes, (**c**) alternate splicing of exons resulting in physiological and venom toxins encoded by the same genes, (**d**) duplication of an ancestral domain resulting in multidomain products or multiple single domain products, and (**e**) consecutive loss of domain from duplicate multidomain precursor genes produce multiple related venom toxins. *Circles* signify gene duplication events (Reproduced with permission, Casewell et al. 2013)

genes (resulting from gene duplication within a population). Because of the latter, it is possible for divergent evolution to occur between two genes within a species.

In the case of divergent evolution, similarity is due to the common origin, such as divergence from a common ancestral structure or function has not yet completely obscured the underlying similarity. In contrast, convergent evolution arises when there is some sort of ecological or physical drivers toward a similar solution, even though the structure or function has arisen independently, such as different characters converging on a common, similar solution from different points of origin. This includes analogous structures. Concerted evolution was originally proposed to explain a high degree of sequence similarity among member genes of a multigene family (Smith 1973). Two models were suggested for the concerted mechanism. In the first model, interlocus recombination mechanism was suggested, which generates new duplicate genes and delete some extant duplicate genes. In the second model, interlocus gene conversion or recombination is regarded as a mechanism of increasing genetic diversity (polymorphism) for a locus or a set of loci by introducing new variants from different loci. Therefore, all loci are assumed to be polymorphic, and polymorphism at different loci evolves in unison as a result of concerted evo-

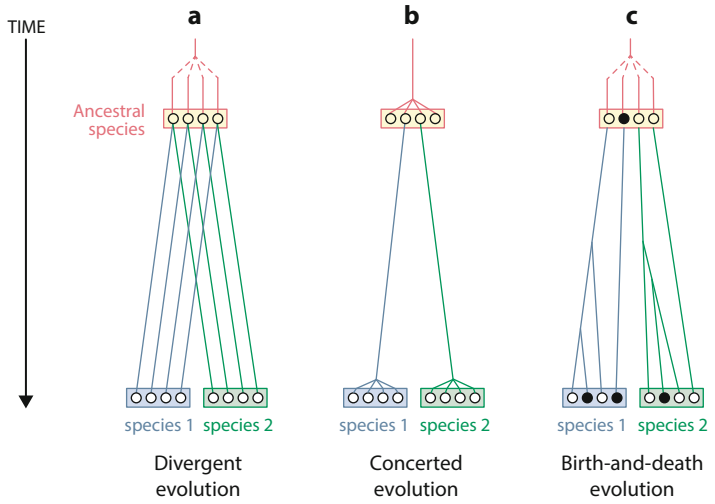

Fig. 5.5 Schematic representation of Divergent, Concerted and Birth and Death Model of evolution. *Open circles* stand for functional genes and *closed circles* for pseudogenes (Reproduced with permission, Nei and Rooney 2005)

lution. In the "birth and death model", duplicate genes are produced by various mechanisms and some of the duplicate genes diverge functionally but others become pseudo-genes owing to deleterious mutations or are deleted from the genomes (Fig. 5.5).

Many venom toxins are thought to evolve via the "birth and death" process (Nei et al. 1997) of gene evolution (Figs. 5.5 and 5.6). In this process, a gene encoding normal physiological genes is duplicated and duplicated gene is expressed in the venom gland (Kordis and Gubensek 2000). Proteins expressed by these genes expressed physiological proteins and exhibit diverse ancestral activities in different cells (Fry et al. 2012; Fry 2005). Once a particular gene is incorporated the gene duplication introduces various functional activities and potencies to the toxin genes (Fry et al. 2003; Weinberger et al. 2011). However, as per earlier theory incorporation of physiological genes to toxin genes is not a one-way process. Toxins can be 'reverse recruited' from the venom gland for a role in physiological processes (Fig. 5.6).

5.6 Mechanism of Evolution of Bacterial Toxins

Evolution of bacterial toxin must comprise of two elements: (a) evolution of a functional progenitor, and (b) adaptation of the gene within the different generation/strains/species. The evolutionary arms race is very apparent in the case of bacteria which build a bewildering variety of toxic peptides and protein to deploy against

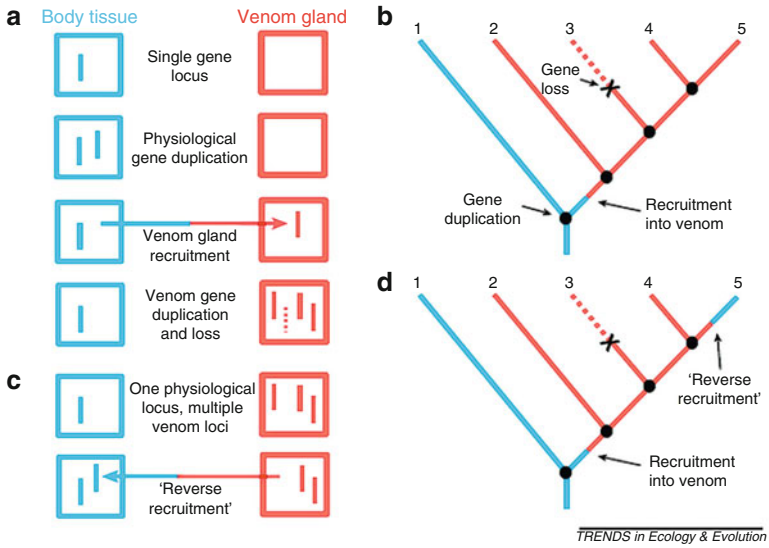

Fig. 5.6 The recruitment and evolution of toxin families under the 'birth and death' model. (**a**) and (**b**) schematics and gene tree of gene duplication and incorporation into the venom gland. (**c**) and (**d**) schematics and gene tree of reverse recruitment (Reproduced with permission, Casewell et al. 2013)

their neighbors or host. Diversity in bacterial toxins, due to evolutionary arms race, makes it difficult to form any hard and fast rule for their evolution. Although some toxin genes are chromosomal, in most cases these are attached with mobile genetic elements such as plasmids, phages, and transposons. For evolution, bacteria must conserve their genetic material and should be able to pass this from one generation to another. Bacteria survive in the changing environmental condition and adapt to the new environment. For this genetic materials need to be flexible. Here genetic materials mean genome plus proteome. Bacteria use point mutations, recombination between homologous DNA sites, the action of transposable genetic elements, and lateral gene transfer as possible mechanisms by which bacteria can have genetic flexibility. Bacteria are able to lose or gain of DNA during adaptation.

Point mutations are considered as driving forces in the slow evolutionary process. During replication, point mutations can be generated by slipped-strand mispairing, resulting in expression or non-expression of particular genes. Regulatory genes which control coordinated gene expression under changing environmental conditions have also been found to be a subject of point mutations or deletion. Random mutagenesis offers bacteria a strong advantage under a selection pressure, which enables bacteria to become more pathogenic without the acquisition of additional genes (Sokurenko et al. 1999). Point mutations in genes contribute to virulence or resistance lead to new variants with new properties. Pathogenicity of bacteria involves two mechanisms: short time interval mechanism (microevolution), and long time interval mechanism (macroevolution).

Lateral gene transfer also plays a significant role in the evolution of bacterial genomes and provides them with a ready to use a novel gene pool. This mechanism is responsible for variation among strains. Documented evidence shows that laterally acquired genes can transform an otherwise avirulent bacteria into a virulent form (Wren 2000; Hacker and Kaper 2000), antibiotic resistance (Alcaine et al. 2005), metabolic diversity (Sullivan and Ronson 1998; Pal et al. 2005), and/or confer an ability to develop a new strain (Campanaro et al. 2005).

There is several nucleotide sequence heterogeneity in the structural genes for some bacterial toxins, such as the heat-labile and heat-stable enterotoxins of *E. coli*. On the other hand, toxins such as botulinum neurotoxins, there are several different antigenic types with similar biological function. Bacterial toxins, in general, have no known role in the physiology of the host bacteria. Bacterial toxins are recognized by their pathogenicity to the eukaryotic organisms. These toxins may have evolved through one of the following processes. First, structural genes for the toxin may be evolved from ancestral genes of eukaryotic proteins which are introduced into prokaryotes and become part of their genes. Second, evolution in prokaryotes may have resulted in the evolution of endogeneous proteins that recognized specific targets in eukaryotes. Finally, due to lateral gene transfer from bacteria and venomous animal (Cordes and Binford 2006).

The diversity of domains, fold composition and extreme sequence divergence are common features of bacterial toxins (Maagd et al. 2003). "Evolutionary arms race" leads to rapid sequence evolution and diversity. This is why it is common to find weak sequence similarity in toxin genes with considerable structural rearrangements (for example, botulinum toxin A (PDB ID: 3BTA) and botulinum toxin E (PDB ID: 3FFZ)). Although, the evolutionary origin and profile of bacterial toxins are not very clear, enough information is available to speculate on a few important possibilities. (a) Surface components of bacterial cells that help in colonization, which is very specific for particular animal host, (b) most bacterial proteins are multifunctional, (c) bacterial proteins are made of different domains, (d) structural similarities among different domains are very less, (e) variation in cell surface receptors, binding, cellular targets are common characteristics of bacterial toxins. These differences could contribute to species-specific susceptibility or resistance and also to the evolution of these toxins.

5.7 Co-evolution

As mentioned earlier, antagonism is the heart of rapid evolutionary change, and life evolves with increasingly complex traits. How do these complexities emerge? Is natural selection is responsible for this or it is just an artifact of adaptation? One of the suggested ideas to answer this question is "coevolution". Several authors have proposed that coevolutionary "arms race" lead to increased complexity. This results in adaptation and counter-adaptations, which ultimately favors complex traits.

The virus, bacteria, parasites, and other microbial pathogens are the foreign invaders that cause illness and death. But most of the pathogens have evolved through a close mutual interaction with their hosts. Because of this interaction, both species develop some unique and specific process either to defend or invade. On one hand, host cells use plasma membrane, antibodies, cytokines, phagocytosis, lysosomes, reactive oxygen species (ROS), low pHs, protease and other factors to prevent invasion by foreign substances, pathogens have developed effective strategies for host defense systems on the other. Pathogen strategies include producing proteins to antagonize ROS effects, inhibit phagosome maturation, block phagosome-lysosome fusion, or produce toxins to disarm the host cell defense.

In general, it is believed that bacterial pathogens have evolved over a long period of time that allows a large number of random interactions of pathogens and host species, possibly leading to coevolution. Further, as suggested earlier, lateral gene transfer could be one of the possible mechanisms of co-evolution? This mechanism is possible when there is some kind of interaction with two kingdoms or species. In general, LGT involves the transfer between prokaryotes. Although examples of LGT involving a transfer between eukaryotes and bacteria are rare, spider toxin sphingomyelinase D (SMaseD) is found to have evidence of this transfer. Notable, SMaseD activity has not been found elsewhere in any other organism except bacteria (Corynebacteria; Soucek et al. 1967; Fig. 5.7). SMaseD of spider and bacteria are similar in size, isoelectric point, enzyme kinetic and substrate specificity (Bernheimer et al. 1985; Truett and King 1993; Van Meeteren et al. 2004). These similarities indicate their distant homology.

In another example, Ambrose et al. (2014) reported LGT of a bacterial toxin gene into a fungus. The bacterial Mcf proteins are considered critical components of the overall insect toxicity of bacteria. Its acquisition by ancestral Epichole fungal endophytes of grass was suggested through LGT (Ambrose et al. 2014). Newly acquired genes, through LGT, can also confer a new capability to the recipient species (Acuna et al. 2012). The exact mechanism of co-evolution is not sure, but it certainly helps in evolution of both the pathogen and the host. Zaman and colleagues demonstrated through their computational experiment that coevolution drives complexity of host (Zaman et al. 2014). Organism evolution is a complex process, requiring interactions of not only environment and biological processes, but also molecular and submolecular processes which provide the chemical and physical basis for biological adaptation and function.

5.8 Structural Diversity and Its Connection with Evolution

Evolution of protein involves a compromise between rigidity and flexibility. Both these factors allow various structural and functional combinations. Internal motion and dynamics are increasingly been recognized as an essential requirement for protein function. One of the naturally existed conformation of the protein, the molten globule state, provides a good model as it has less rigidity and more flexibility than the native

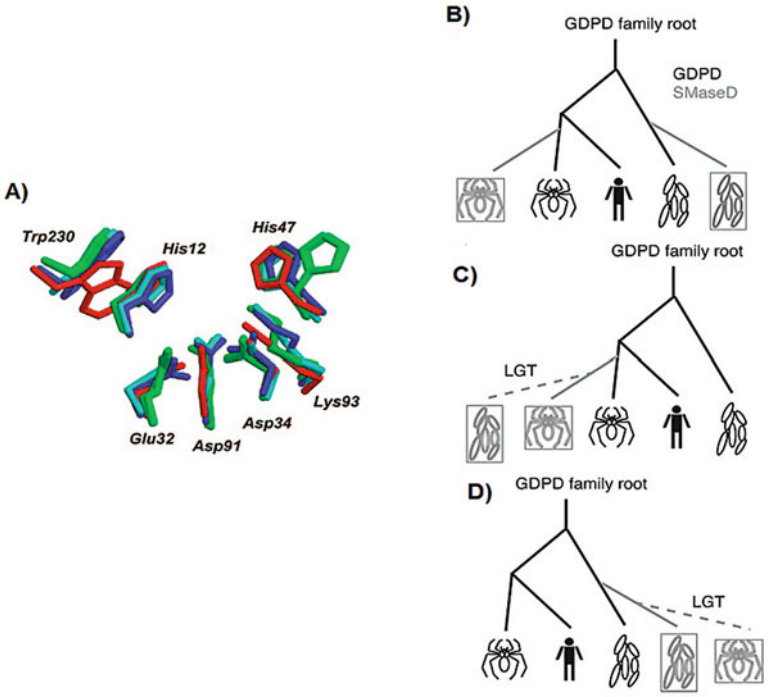

Fig. 5.7 (**a**) Active site superposition of E. coli cytoplasmic and periplasmic GDPDs (Glycerophosphoryl diester phosphodiesterase) (1T8Q, 1YDY_A and 1YDY_B, *all green*), Thermotoga maritime GDPD (1O1Z, *cyan*), Thermus thermophilus GDPD (1V8E, *blue*) and Loxosceled laeta SMase I (1XX1_A, *red*). For 1T8Q and 1XX1, all chains in the asymmetric unit showed similar rotamers for all residues. For 1YDY, one of the two chains showed a conformation for His 47 (SMase I numbering) similar to that of the E. coli cytoplasmic GDPD (1T8Q), while the other showed a conformation similar to that of the other GDPDs and the L. laeta SMase I. SMase originate from divergence of GDPDs. Phylogenetic scenarios representing hypothetical mechanisms of origin of SMaseD activity (*gray*) from within the glycerophosphoryl diester phosphodiesterase (GDPD) family. (**b**) Independent duplication and functional divergence in both bacteria and spiders. (**c**) Duplication and functional divergence in spiders followed by LGT to bacteria. (**d**) Duplication and functional divergence in bacteria followed by LGT to spiders. Variants of (**b**) and (**c**) are also possible in which acquisition of SMaseD activity occurs independently in bacteria and spiders following an LGT event (Reproduced with permission, Cordes and Binford 2006)

state. It is now believed that molten globules retain their functional activities, even though the other sites are fluctuating or in misfolded conformations, for example botulinum toxins (Chaps. 3 and 4). Flexibility gives several advantages. (i) A large number of variations are possible without much effect on the structure or function. Variations are presumable either neutral or compensated by proper mutations. (ii) The highly flexible regions, in turn highly evolving, serves as substrate recognition sites. This also pointed out that conformational adaptability is also important in physicochemical specificity (Luque and Freire 2000; Dobbins et al. 2008). (iii) If the interaction (for example, binding) surface area is large, then because of flexibility

pre-organization of residues with restricted and higher mobility is possible. This helps in stabilizing bound conformer, which is able to compensate unfavorable entropic factors with favorable enthalpic elements. (iv) As a result of flexibility, it is easier for species or molecules to make a balance between physical adaptability, chemical specificity, and stability in the hostile environment. (v) Intrinsic flexibility is intimately related to the formation of protein complexes (Marsh and Teichmann 2012).

Protein toxins are either multi-domain or complex proteins. Is there any role of flexibility in the assembly of these toxins? Proteins which form a larger surface area of interactions or larger inter-subunit interfaces have less surface area available to bury intra-molecularly are more likely to be flexible in isolation. In addition, subunit flexibility should be strongly associated with the formation of interaction within the complex. Is there any relationship between flexibility and symmetry? It is easy to imagine that flexibility would be beneficial in case of the asymmetric heterologous molecules. Dimeric isologous interfaces involve self-complementary surfaces. In other words, if there is conformational change on one side of a dimeric isologous molecule, then there must be equivalent conformational change on the other side. This requirement makes intrinsic flexibility, less advantageous for these molecules. So it is easy to hypothesize that asymmetric surfaces require larger flexibility than the homomers or dimers. However, flexibility has a significant correlation with the secondary structure; α proteins tend to be more flexible than β-proteins (Marsh 2013). A very good example to support above arguments is botulinum neurotoxin. Botulinum neurotoxin (BoNT) is a protein complex (LL complex is ~900 kDa, L complex is ~600 kDa and M complex is ~300 kDa) in which toxin molecule is surrounded by several non-toxic proteins. These surrounded proteins have a role in stability, transport and functional activity of the toxin. Assembly of these proteins requires flexibility of toxins as well as subunits. This flexibility is one of the reasons for having several serotypes and subtypes for this protein family, and assists in trafficking (through the harsh environment of gut and blood) and substrate specificity. Two members of this family, BoNT/A and BoNT/E, are optimally active in their native molten globule conformation (Cai and Singh 2001; Kukreja et al. 2010). A toxin molecule of BoNT complex utilizes most of its surface for interactions with NBP (Neurotoxin Binding Protein). This may be the reason for the existence of flexible structure in isolation and for the existence of a biologically active molten globule conformation (as mentioned earlier in this paragraph). The structure of toxin alone is very interesting which consists of a binding domain (β-sheet protein), a translocation domain (α-helix protein), and a catalytic domain (α/β protein). This unique structure is self-explanatory; for binding molecule needs some rigidity which is provided by binding domain secondary structural organization (β-sheets), for translocation molecule need fluid structure (α-helix), and for catalysis needs binding as well as function (α/β). This structural organization and exquisite coordination between different parts of this molecule makes this the most potent molecule.

Contrary to the above view, cry proteins (Maagd et al. 2003), three-finger and spider toxins are mainly β-sheets proteins with structural and functional divergence (described above). The structural and functional divergence, and inherent flexibility

of these proteins is the result of the differential organization of β-sheets and loops. So, it may not be appropriate to directly relate flexibility of protein to secondary structure. The flexibility of any protein is a fine play between the organization of secondary structure, the arrangement of loops, and overall tertiary structure. This in turn helps molecules to survive, adapt, and evolve.

5.9 Molecular Flexibility and Functional Implications

As mentioned, flexibility could be an important parameter for functional implications. Best example for this is the most potent molecule known: Botulinum Neurotoxins (BoNTs). The protein is produced by one of the oldest organisms (anaerobic organism estimated to be ~2 billion years old), and seem to have evolved into at least seven known serotypes and dozens of subtypes which target an entire process of exocytosis, rather than an individual molecule as one would expect for the evolution of a group proteins with very strong specificity. BoNTs, though endopeptidase of thermolysin type, are extremely specific, and do not proteolyze any other than three SNARE proteins involved in the exocytosis process. This endopeptidase has very low turnover compared to other enzymes, but because of its high specificity, selectivity, physiological response, and exceptionally high stability inside cell, is extremely effective. It provides a rich and a unique biological system to understand and describe the versatility of nature.

The protein smartly avoids its destruction in the endosome by avoiding lysosomal attack. In addition, it is now well known that it can survive intracellularly for several months, a feature unknown for any other exogenous protein. Structural and host response features are beginning to provide clues, albeit at times contradictory, to the mystery of this protein.

The role of its dynamic tertiary structure is significant in determining its specificity towards its substrates, SNARE proteins (Kumar et al. 2014). To the best of our knowledge BoNT is the only example of fully enzymatically active molten globule (MG). Our latest experiments have revealed the following: (a) The unfolding pattern BoNT/A endopetidase domain has a very unique intermediate state showing dual unfolding. We believe this protein has inherent flexibility which is due to its backbone vibrations leading to a set of folding domains not predictable from the crystal structure (Kumar et al. 2013a, 2014). (b) The substrate peptide undergoes preinteraction molecular motion (as indicated by decreased fluorescence anisotropy), before its binding to the enzyme and subsequent cleavage (unpublished data). (c) Among the three BoNT serotypes (BoNT/A, B, and E) examined for structural flexibility, BoNT/E provides the most flexible structure, which matches with evolutionary dendrogram based on BoNT sequence analysis showing BoNT/E to be evolutionarily latest BoNT serotype (Kumar et al. 2013b; Chang 2011). Such findings may be revolutionary in suggesting molecular flexibility and dynamics as criteria for molecular evolution. Interestingly, light chains (LCs; A, B, and E) of these serotypes behave differently in solution, both structurally and functionally (Kumar et al. 2014).

Furthermore, catalytic domain of this protein (LC) has Zn^{2+} in its active site and is believed to have both structural and functional roles (Li and Singh 2000). This may provide additional evidence that the Zn^{2+} binding in the active site may influence the dynamics of the protein.

Other steps involved in BoNT's mode of action, in addition to the enzyme activity, also seem to incorporate protein flexibility. In one of the models describing BoNT translocation into cytosol from endosomes, the LC undergoes molten globule transition at low pH. Both the BoNT/A toxin and the catalytic domain were shown to have molten globule states. Structural changes observed for BoNT/A LC at low pH leading to elongated molten globule (Cai et al. 2006) is believed to facilitate translocation of BoNT/A LC into the cytosol, where it renatures and regains its endopeptidase activity. The flexibility of molten globule may not only be relevant to the protein chemistry involved in the functioning of the BoNT molecule, but may also play a major role in the survival of the protein molecule intracellularly for months. In substrate-ligand binding, anchoring and alignment regions involve hydrophobic, salt bridge and hydrogen bond interactions indicating common evolutionary origin with substrate.

There are reports of other proteins to involve molten globule state for insertion and translocation (see Chap. 4). Structural analysis of physiologically relevant intermediates and molten globule states of BoNT will provide valuable information on the questions related to this molecule as well as about the critical role of biophysical forces and processes in the protein chemistry.

5.10 Summary

Toxins are some of the oldest and most evolve molecules on this Earth. Structural and functional diversity, co-evolution, and complex symbiotic relationship with their host hinder experimentalists/theorists in developing an evolutionary theory for these molecules. But these unique molecules provide an important natural model, which can provide better mechanistics of evolution at the molecular level. These molecules are perfect candidates to study predator-prey relationship, mutuality, parasitic, commensalism, and mimicry. Several models are suggested in the literature and every model fundamentally indicates towards organism or molecular flexibility. Molecular flexibility allows toxins to adapt and diversify. For example, alteration of protein surface allows the molecule to have the same fold with different functions (Sect. 4.5). This is followed by rapid transfer of molecular properties from one organism to another to facilitate co-evolution. On the other side, molecular flexibility also allows organization of different domains required by toxin molecules to counter hostile environment of the target. Due to co-evolution toxin molecules are able to find a new target, function and establish the new relationship with their host/prey. Finally, because of the newly established relationship, due to co-evolution, toxin molecules redefine their role in the evolution.

References

Acuna R, Padilla BE, Claudia PFR et al (2012) Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. PNAS 109:4197–4202

Alcaine SD, Sukhnanand SS, Warnick LD et al (2005) Ceftiofur-resistant Salmonella strains isolated from dairy farms represent multiple widely distributed subtypes that evolved by independent horizontal gene transfer. Antimicrob Agents Chemother 49:4061–4067

Ambrose KV, Koppenho AM, Belanger FC (2014) Horizontal gene transfer of a bacterial insect toxin gene into the Epichloe fungal symbionts of grasses. Sci Rep 4:1–8

Bernheimer AW, Campbell BJ, Forrester LJ (1985) Comparative toxinology of Loxosceles reclusa and Corynebacterium pseudotuberculosis. Science 228:590–591

Bergelson J, Dwyer G, Emerson JJ (2001) Models and data on plant-enemy coevolution. Annu Rev Genet 35:469–499

Binford GJ, Bodner MR, Cordes MH et al (2009) Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase d in sicariid spider venoms. Mol Biol Evol 26:547–566

Cai S, Singh BR (2001) Role of the disulfide cleavage induced molten globule state of type A botulinum neurotoxin in its endopeptidase activity. Biochemistry 40:15327–15333

Cai S, Kukreja R, Shoesmith S et al (2006) Botulinum neurotoxin light chain refolds at endosomal pH for its translocation. Protein J 25:455–462

Campanaro S, Vezzi A, Vitulo N et al (2005) Laterally transferred elements and high pressure adaptation in Photobacterium profundum strains. BMC Genomics 6:122

Casewell NR, Wuster W, Vonk FJ et al (2013) Complex cocktails: the evolutionary novelty of venoms. Trends Ecol Evol 28:219–229

Chang TW (2011) Sequence analysis and novel antidotes development of botulinum neurotoxin. PhD Thesis, University of Massachusetts, Dartmouth.

Chang D, Duda TF Jr (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. Mol Biol Evol 29:2019–2029

Chippaux JP, Williams V, White J (1991) Snake venom variability: method of study, results and interpretation. Toxicon 29:1279–1303

Clay K, Kover PX (1996) The red queen hypothesis and plant/pathogen interactions. Annu Rev Phytopathol 34:29–50

Cordes MHJ, Binford GJ (2006) Lateral gene transfer of a dermonecrotic toxin between spiders and bacteria. Bioinformatics 22:264–268

Daltry JC, Wuster W, Thorpe RS (1996) Diet and snake venom evolution. Nature 379:537–540

Deshimaru M, Ogawa T, Nakashima K et al (1996) Accelerated evolution of crotalinae snake venom glan serine proteinases. FEBS Lett 397:83–88

Dobbins SE, Lesk VI, Stenberg MJ (2008) Insight into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. PNAS 105:10390–10395

Duda TF Jr, Palumbi SR (1999) Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod Conus. PNAS 96:6820–6823

Escoubus P, King GF (2009) Venomics as a drug discovery platform. Expert Rev Proteomics 6:221–224

Fry BG (2005) From genome to venome: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome Res 15:403–420

Fry BG, Wüster W, Kini RM et al (2003) Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J Mol Evol 57:110–129

Fry BG, Roelants K, Champagne DE et al (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet 10:483–511

Fry BG, Casewell NR, Wuster W et al (2012) The structural and functional diversification of Toxifera reptile venom system. Toxicon 60:434–448

Gong N, Armugam A, Jeyaseelan K (2000) Molecular cloning, characterization and evolution of the gene encoding a new group of short chain α-neurotoxins in an Australian elapid, Pseudonaja textilis. FEBS Lett 473:303–310
Hacker J, Kaper JB (2000) Pathgenicity islands and the evolution of microbes. Annu Rev Microbiol 54:641–679
Harris JB (1998) Phospholipase A2 that show neurotoxic activity. In: Bailey GS (ed) Enzymes from snake venoms. Fort Collins, Alaken, pp 425–449
Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. Proc Biol Sci 256:119–124
Hughes AL (2000) Adaptive evolution of genes and genomes. Oxford University Press, New York
Hughes AL, Nei M (1998) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 355:167–170
Kini RM (1997) Phospholipase A2: a complex multifunctional protein puzzle. In: Kini RM (ed) Venom phospholipase A2 enzymes: structure, function and mechanism. Wiley, Chichester, pp 1–28
Kini RM, Chan YM (1999) Accelerated evolution and molecular surface of venom phospholipase A2 enzyme. J Mol Evol 48:125–132
Kordis D, Gubensek F (2000) Adaptive evolution of animal toxin multigene families. Gene 261:43–52
Kukreja RV, Sharma SK, Singh BR (2010) Molecular basis of activation of endopeptidase activity of botulinum neurotoxin type E. Biochemistry 49:2510–2519
Kumar R, Kukreja RV, Li L et al (2013a) Botulinum neurotoxin: unique folding of enzyme domain of the most-poisonous poison. J Biomol Struct Dyn 32:804–815
Kumar R, Chang TW, Singh BR (2013b) Evolutionary traits of toxins. Handbook of toxinology. In: Gopalakrishkone P (ed) Biological toxins and bioterrorism, vol 1, chapter 23. Springer, Dordrecht, Heidelberg, New York, London, p 527–557
Kumar R, Kukreja RV, Cai S et al (2014) Differential role of molten globule and protein folding in distinguishing unique features of botulinum neurotoxin. Biochim Biophys Acta 1844:1145–1152
Li L, Singh BR (2000) Role of zinc binding in type A botulinum neurotoxin light chain's toxic structure. Biochemistry 39:10581–10586
Luque I, Freire E (2000) Structural stability of binding sites: consequences for binding affinity and allosteric effects. Proteins Suppl 4:63–71
Maagd RA, Bravo A, Berry C et al (2003) Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu Rev Genet 37:409–433
Marsh JA (2013) Buried and accessible surface area control intrinsic protein flexibility. J Mol Biol 425:3250–3263
Marsh JA, Teichmann SA (2012) Protein flexibility facilitates quaternary structure assembly and evolution. PLoS Biol 12:4197–4202
Mebs D (2001) Toxicity in animals. Trends in evolution? Toxicon 39:87–96
Menez A (1998) Functional architectures of animal toxins: A clue to drug design. Toxicon, 36, 1557–1572
Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39:121–152
Nei M, Gu X, Sitnikova T (1997) Evolution by the birth and death process in multigene families of the vertebrate immune system. PNAS 94:7799–7806
Nobuhisa I, Deshimaru M, Chijiwa T et al (1997) Structures of genes encoding phospholipase A2 inhibitors from the serum of Trimeresurus Flavouridis snake. Gene 191:31–37
Norton RS, Pallaghy PK (1998) The cystine knot structure of ion channel toxins and related polypeptides. Toxicon 36:1573–1583
Ohno M, Menez R, Ogawa T et al (1998) Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? Prog Nucleic Acid Res Mol Biol 59:307–364
Olivera BM (2002) Conus venom peptides: reflections from the biology of clades and species. Annu Rev Ecol Syst 33:25–47

Pal C, Papp B, Lercher MJ (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat Genet 37:1372–1375
Rost B (1997) Protein structures sustain evolutionary drift. Fold Des 2:S19–S24
Sandegren L, Andersson DI (2009) Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat Rev Microbiol 7:578–588
Smith GP (1973) Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp Quant Biol 3:507–513.
Smith WL, Wheeler WC (2006) Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms. J Hered 97:206–217
Smith JJ, Hill JM, Little MJ et al (2011) Unique scorpion toxin with a putative ancestral fold provides insight into evolution of the inhibitor cystine knot motif. PNAS 108:10478–10483
Sokurenko EV, Hasty DL, Dykhuizen DE (1999) Pathoadaptive mutations: gene loss and variations in bacterial pathogens. Trends Microbiol 7:191–195
Soucek A, Michalec C, Souckova A (1967) Enzymatic hydrolysis of spingomyleins by a toxin of Corynebacterium ovis. BBA Lipid Lipid Metab 144:180–182
Stock EP, Stock AM, Mottonen JM (1990) Signal transduction in bacteria. Nature 344:395–400
Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500 kb symbiosis island that integrates into a phe-tRNA gene. PNAS 95:5145–5149
Sunagar K, Jackson TNW, Undheim EAB et al (2013) Three-fingered RAVERs: rapid accumulation of variations in exposed residues of snake venom toxins. Toxins (Basel) 5:2172–2208
Takasaki C, Yutani F, Kajiyashiki T (1990a) Amino acid sequence of eight phospholipase A2 from the venom of Australian king brown snake. Toxicon 28:329–339
Takasaki C, Sugama A, Yanagita A et al (1990b) Effects of chemical modifications of Pa-11, a phospholipase A2 from the venom of Australian king brown snake (Pseudechis australia), on its biological activities. Toxicon 28:107–117
Toft C, Andersson SGE (2010) Evolutionary microbial genomics: insight into bacterial host adaptation. Nature 11:465–475
Truett AP III, King JLE (1993) Sphingomyelinase D: a pathogenic agent produced by bacteria and arthropods. Adv Lipid Res 26:275–291
Van Meeteren LA, Frederiks F, Giepmans BN et al (2004) Spider and bacterial Sphingomyelinase D target cellular lysophosphatidic acid receptors by hydrolyzing lysophosphatidylcholine. J Biol Chem 279:10833–10836
Weinberger H et al (2011) Positions under positive selection—key for selectivity and potency of scorpion α-toxins. Mol Biol Evol 27:1025–1034
Wren BW (2000) Microbial genome analysis: insight into virulence, host adaptation and evolution. Nat Rev Genet 1:30–39
Zaman L, Meyer JR, Devangam S et al (2014) Coevolution drives the emergence of complex traita and promotes evolvability. PLoS Biol 12:1–9

Epilogue

Chose a job you love, and you will never work a day in your life.

Confucius

This book is a concise effort of presenting toxin as a model for examining various aspects of protein chemistry. Toxin family offers a unique group of molecules which is structurally and functionally diverse. In addition to this, they also add another unique dimension to study protein structure and function.

Toxin as Molten Globule

After the discovery of protein folding intermediates, the study of intermediates become important in Protein Folding Problem (PFP). One of the folding intermediates is molten globule. This state is very important in the structure and function of the toxins. Disorder state is often transiently populated during protein folding. Generally speaking, highly dynamic protein states do not exhibit catalytic activity. For example, I116G/L186G of adenylate kinase and S45G/P55N of RNase T1 in the partial disordered state (higher dynamic state) is 5 and 40 % active, respectively. In contrast, acylphosphatase in its partially disordered state is 79 % active. Also, in molten globule state DHFR (Di-Hydrofolate Reductase) and human topoisomerase I are 5 and 11 % active, respectively. Whereas mutated SNase (F34W/W104F) in molten globule conformation is ~87 % active. Until now none of the native enzymes in molten globule conformation is optimally active, except botulinum toxin. Several toxins have been shown to acquire the functional state by transforming into a molten globule conformation. Not only for catalytic activity, molten globule conformation of toxins is also suggested for membrane insertion, channel formation, translocation, chaperon activity, binding, ligand transfer and other activities. Examples are colicins, diphtheria toxin, botulinum toxin, anthrax toxin, etc.

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

Toxin as IDPs

Toxins have a great amount of pressure to escape from the damage caused by the target environment. Toxins mastered this by getting inherent flexibility which facilitates them to exist in several functional conformations. Flexibility gives toxins the required freedom which is advantageous in interaction, binding, selectivity, and longevity. According to DisProt database, out of 700 know disordered proteins 14 are toxin or toxin associate proteins. Also BoNT/A, BoNT/B, BoNT/E, tetanus, diphtheria and anthrax lethal factor have significant amount of disorders.

Toxins as Evolutionary Model

In the primordial environment, organisms need molecules which can adapt to changing environment. The existence of toxins as IDPs and molten globules offers two advantages to organisms; (a) better flexibility provides a sophisticated mechanism for many-to-one and one-to-many interaction, and (b) reduce energy cost to store and maintain genomic size (reduce the need to have large gene pool). Structural and functional diversity, possibility of co-evolution with the substrate, and complex symbiotic relationship offer two concepts about these molecules; (a) hinders in developing a consensus conventional evolutionary theory, and (b) makes them a good tool to redefine evolutionary paradigm.

In summary, this book addresses various aspects of protein functions and introduces toxin as a better model to understand molecular basis of their existence and action.

Index

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