Dev Bukhsh Singh
Timir Tripathi *Editors*

Frontiers in Protein Structure, Function, and Dynamics

Springer

Frontiers in Protein Structure, Function, and Dynamics

Dev Bukhsh Singh • Timir Tripathi
Editors

Frontiers in Protein Structure, Function, and Dynamics

Springer

Editors
Dev Bukhsh Singh
Institute of Biosciences and Biotechnology
Chhatrapati Shahu Ji Maharaj University
Kanpur, Uttar Pradesh, India

Timir Tripathi
Department of Biochemistry
North Eastern Hill University
Shillong, Meghalaya, India

ISBN 978-981-15-5529-9 ISBN 978-981-15-5530-5 (eBook)
https://doi.org/10.1007/978-981-15-5530-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Proteins are highly complex biomolecules. They are the most diverse group of biologically important substances and are often considered to be the central compound necessary for life. They play major roles in cellular metabolism, defense, communication, transport, storage, and recognition; all the processes are required for the structure, function, and regulation of the body's tissues and organs. The collection of proteins within a cell determines its health and function. Proteins interact with other biomolecules to perform biological functions. Accordingly, protein science has been a center of basic research as well as applied to areas such as biotechnology, medicine, and agriculture.

The structure of a protein is defined by its amino acid sequence. The central dogma of structural biology states that a folded three-dimensional protein structure is essential for its biological function. However, the view of protein structures has recently changed from native structures being considered to be a single, rigid, static entities into dynamic entities with internal motions where conformational ensembles coexist. These motions and the resulting conformational dynamics play an essential role in protein functions. Therefore, full understanding of the protein structure–function relationship requires knowledge of protein dynamics in correlation with biological systems. Proteins typically do not function alone, but in physical or functional interaction with other proteins, forming macromolecular complexes. The dynamic nature of proteins is often altered upon the binding of ligands, proteins, or nucleic acids. Several covalent and non-covalent interactions play a critical role in maintaining the structural integrity of the protein. Hence, in order to understand the nature of protein, it is important to understand and quantify all possible interactions that control protein structure and dynamics that regulate its functions.

In the last few decades, experimental techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy have made substantial progress. Going beyond the static picture of single protein structures has been more challenging. Nowadays, a number of techniques such as NMR relaxation, fluorescence spectroscopy, and single molecule methods have been developed which provide information on the conformational flexibility and dynamics of proteins. However, computational simulation methods such as molecular dynamics (MD) currently provide the routine means to obtain information on protein dynamics at an atomic level on timescales of nano- to microseconds.

This book provides a compendium of important topics on advanced research in the field of protein science. Each chapter addresses specific aspects of proteins with a definitive theme designed to enhance the reader's understanding of protein isolation to structure, function, and dynamics. The latest developments in the field have been discussed and key concepts introduced in a simple yet comprehensive way to ensure that readers can grasp the essentials. This book will be an invaluable resource that would be very useful to undergraduates, graduate students, postdoctoral researchers, and instructors involved in biophysics, structural biology, or bioinformatics courses or research on protein structure–function relationships.

Chapter 1 provides an overview of protein extraction, purification, estimation, storage, and dynamics, which is the first step in understanding protein structure, function, and dynamics. Chapters 2 and 3 cover the major aspects of protein purification techniques, protein characterization methods, protein structure determination approaches, and the importance of molecular interactions in protein stability. Experimental protein structure determination methods including nuclear magnetic resonance, X-ray crystallography, and cryo-electron microscopy are discussed. Computational strategies to predict protein structure and conformation are also considered. Chapters 4 and 5 discuss the use of group-specific reagents used in the identification of active-site functional groups of enzymes. The kinetics of inactivation, effects of protectants, kinetic parameters, and spectral analysis are also elaborated.

Understanding of protein–protein interactions and complex formation allows analysis of molecular functions. Protein–protein interactions can be studied through various physical, molecular, and genetic methods. The goal of Chap. 6 is to present various methods from wet and dry labs to rationalize tools for the investigation of protein interactions of various nature and strength. Chapter 7 describes the importance of protein–ligand binding, the thermodynamic parameters describing the binding and the methods to measure these parameters. Chapter 8 describes the importance of water molecules maintaining protein structural stability, folding, competition with the ligand, and other biomolecular processes. Chapter 9 provides comprehensive information on various approaches and packages used in MD simulations of biomolecules. Chapter 10 provides fundamental understanding of the single-molecule fluorescence techniques and their applications in studying protein folding and aggregate formation in disease conditions. Separate sections in this chapter discuss the use of these techniques on the study of both folded as well as intrinsically disordered proteins.

Misfolding and aggregation of proteins and their accumulation as fibrils are hallmarks of protein misfolding disorders and several neurodegenerative diseases. Chapter 11 discusses how protein quality control fails during various protein misfolding diseases. The mechanism of protein misfolding and the roles of molecular chaperones during folding are also discussed. Chapter 12 discusses the strategies that exist in protein sequences to scuffle their uncontrolled aggregation. Using the example of insulin aggregation in diabetes, the chapter considers how intracellular protein deposits can be modulated for therapeutic purposes. Chapter 13 provides information on the structure and function of different types of muscarinic

acetylcholine receptors and their target class of antimuscarinic drugs, mechanism, epidemiology, side effects, antagonistic effect, and clinical significance.

Chapter 14 outlines the current state of understanding of dopamine beta-hydroxylase, describing its therapeutic importance involving the heart and the brain, with structure-based inhibitors expected to combat hypertension and cocaine addiction specifically. Chapter 15 focuses on the structural and molecular basis of force generation that leads to the directional motility in molecular motors, taking an example of kinesin. Recent findings on the bidirectional movement of kinesin-5 motors and several structural determinants that regulate directional switching are also discussed. Chapter 16 emphasizes on the structure–activity relationship of cathepsins cysteine proteases. It attempts to enhance understanding of cathepsins and presents them as drug targets to control the degradation of extracellular matrix and other components of the disease. Chapter 17 outlines the important structural and functional roles of the iron-binding protein ferritins in the physiology of *Mycobacterium tuberculosis.*

Kanpur, Uttar Pradesh, India — Dev Bukhsh Singh
Shillong, Meghalaya, India — Timir Tripathi

Contents

About the Editors

Dev Bukhsh Singh is an Assistant Professor at the Department of Biotechnology, Chhatrapati Shahu Ji Maharaj University, India. He received his Ph.D. in Biotechnology from Gautam Buddha University, India, after completing his B.Sc. and M.Sc. degrees at the University of Allahabad, Prayagraj, and M.Tech. in Bioinformatics at the Indian Institute of Information Technology, Prayagraj, India. He has been actively involved in teaching and research since 2009, particularly in the areas of molecular modeling, chemoinformatics, inhibitor/drug design, and *in silico* evaluation. He has authored many research and review articles and book chapters in the fields of medicinal research, molecular modeling, drug design, and systems biology. He is a member of various national and international academic and research organizations, and has served as a referee for several international journals published by Springer, Bentham Science, and Taylor & Francis.

Timir Tripathi is a Senior Assistant Professor at the Department of Biochemistry, North-Eastern Hill University, Shillong, India. He obtained his Ph.D. degree from the Central Drug Research Institute, Lucknow in 2010. He was a visiting faculty at ICGEB, New Delhi, India (2011), and KhonKaen University, Thailand (2015). His research interests are focused on understanding the structure-function-folding-misfolding-aggregation-flexibility-dynamics relations in proteins. He has done pioneering work on the protein-substrate interaction & dynamics, and understanding the roles of non-catalytic domains in regulating the catalytic activity of key drug target proteins. His work has provided a platform for the development of novel therapeutic strategies. He has received various awards including the Dr. D.M. Bose Award (2008), DST Fast Track Young Scientist Project Award (2012), DBT Overseas Associateship Award (2012), BRSI- Malviya Memorial Award (2017), ICMR-Shakuntala Amir Chand Prize (2018), and ISCB-Young Scientist Award (2019). In 2019, he was elected as a member of the National Academy of Sciences, India (NASI) and the Royal Society of Biology (RSB), London. He has published more than 70 research papers, reviews, commentaries and viewpoint articles in highly respected international journals. He has handled ten research grants as PI from various national and international funding agencies including the DBT, DST,

SERB, UGC, ICMR, Russian Foundation for Basic Research, and Israel Science Foundation. He serves on the editorial boards of several journals including Scientific Reports (Nature Publishing Group) and Plos One. He is a Management Council Member of the Biotech Research Society of India (BRSI), and of the Bioinformatics and Drug Discovery Society (BIDDS), India.

Protein Purification, Estimation, Storage, and Effect on Structure–Function–Dynamics

1

Awanish Kumar

Abstract

Researchers working on structure, function, and dynamics of proteins are cautious about the process of protein purification, quantification after purification, and the storage. One faces a lot of problems, and there is always an array of confusion as where to start in the selection of an appropriate method for the high production and easy purification of recombinant proteins. The development and availability of many commercial systems have made protein engineering work easier and more widespread for the highly purified protein. After purification, we estimate the protein and need in milligram concentration for further downstream applications. For the quantification of purified protein, I describe some simple and common methods in this chapter. Storage of purified protein is an important issue because concentrated proteins are more unstable to freeze-thaw cycles. After purification, care should be taken for the storage and use of proteins because protein may get degraded or aggregated due to erroneous handling. Checking the stability of the protein after the storage is very necessary from the structure–function–dynamics point of view of protein before further processing. The chapter would be very helpful regarding the basic knowledge and understanding of the proposed title.

Keywords

Protein extraction · Purification · Estimation · Assay storage · Dynamics

A. Kumar (✉)
Department of Biotechnology, National Institute of Technology, Raipur, Chhattisgarh, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_1

1.1 Introduction

Protein is a class of biomolecules, which is present in every living cell and performs an important function. They are known as the building blocks of life. They are found throughout the body of living in the form of callus, cartilage, hair, ligaments, muscles, skin, tendons, etc. Proteins protect the cell and provide a definite structure to the body of a multicellular organism. Antibodies, enzymes, and hormones are made up of proteins. Within the body, protein (hemoglobin, myoglobin, and various lipoproteins) regulates the transport of oxygen and other substances (Urry et al. 2017). They catalyze and regulate cellular metabolism and maintain the body. Proteins are made from amino acids that are synthesized within a living cell by the process of translation. Triplet codon of mRNA codes for an amino acid during the process of translation and many amino acids together form a protein. A dipeptide is formed when the carboxyl group of one amino acid reacts with the amino group of another, creating a peptide bond –C(=O)NH– with a subsequent release of one water molecule. Further, linking of more amino acids can create a peptide chain. A peptide having ten or more amino acids is known as oligopeptide. Polypeptides are chains of ten or more amino acids; polypeptides consisting of more than 50 amino acids are classified as proteins (Lodish et al. 2000).

Proteins are a necessary part of an animal's diet because cellular physiology needs it. If an individual does not eat adequate protein, they can become ill seriously. Certain proteins also have therapeutic value. Protein-based antibiotics, antibodies, and vaccines are also used in the treatment of diseases (Ahuja 2006). The Institute of Medicine (Washington, DC) (2005) recommends that adults get a minimum of 0.8 g of protein for every kg of body weight per day. The Institute also sets a wide range for acceptable protein intake anywhere from 10 to 35% of calories each day (https://www.nap.edu/read/10490/chapter/1). Therapeutic proteins are mostly developed by recombinant protein production and purification. This chapter describes the extraction, purification, estimation, and storage of protein that is very necessary from the structure–function–dynamics viewpoint of protein and further downstreaming studies (Fig. 1.1).

1.2 Protein Purification

To study the functions and dynamics of proteins, we first need to purify it as it would be very difficult to study the proteins if they are present in the mixture. This will cause errors and wrong information about the proteins. Purification is a stepwise process of separating specific proteins (one or more) from cells, tissues, and organs (Berg et al. 2002). Proteins are purified in small to large volumes for various biotechnological applications. According to the US Food and Drug Administration, the degree of purity is an essential factor for injectable proteins (US Food and Drug Administration 1991). In general, protein is worthless, if contaminants are detected in the purified proteins. Specific protein is purified from the pool of protein and they are not only purified away from other proteins but should be free from other

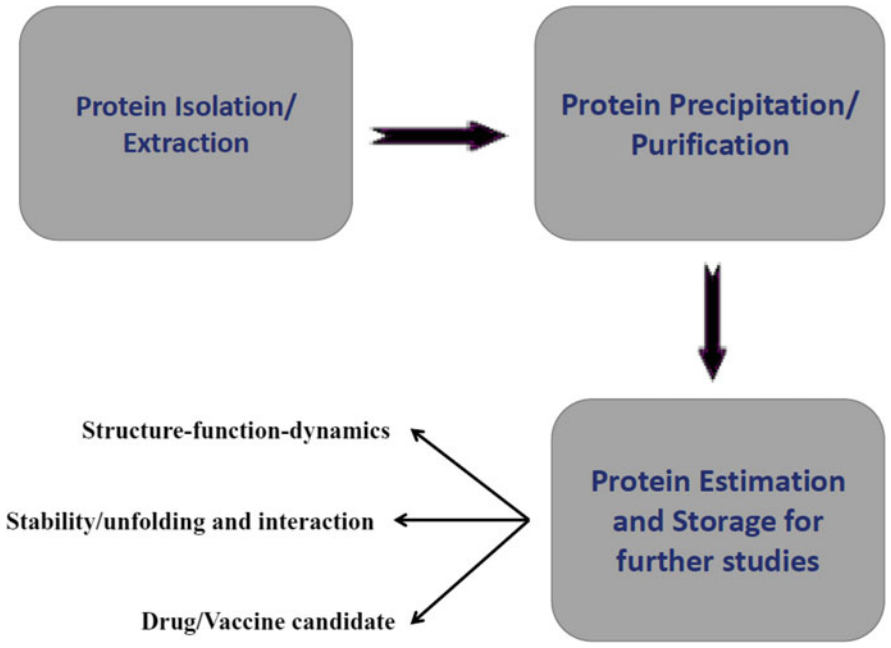


Fig. 1.1 Workflow of protein purification and downstreaming studies

macromolecules (carbohydrates, lipids, nucleic acids) of the sample (Arentson-Lantz et al. 2015). We must develop a specific test for contaminants present in a protein sample if the source material (sample) contains any unusual things. Purified protein must retain its biological activity. The FDA does not certify a procedure of protein production that results in invariably biological activity. The procedure must harvest the same quality and amount of protein in every batch with the same biological activity. This complies with the use of a very vigorous process of protein purification that could be used to purify a biological activity protein (Berg et al. 2002). A method that works nicely in a research lab may fail wretchedly on the floor of protein production, where it must be scaled up and reproduced accurately every day.

The researcher makes use of inherent differences and similarities to purify proteins. The concept of differences is used to purify one protein from another. One protein differs from other in biological activities, charge, hydrophobicity, shape, size, solubility, etc. Non-protein contaminants like carbohydrates, lipids, and nucleic acids also hinder the process of protein purification. All the above differences can be used for the separation of the proteins in a complex solution that is produced by bursting *E. coli* host cells to get a recombinant protein (Gopal and Kumar 2013). We can make the work of protein purification easier by attaching a fusion "tag" to the protein sometimes, but again it is rarely allowed by the FDA for the injectable proteins. Therefore, protein purification should be done smartly where technology and innovation must be used together to make the process efficient and reliable.

Purification of proteins is an art for the best practices in research and industry. There are various techniques and processes involved in the purification process, but before purification extraction of protein from the source is required (Coskun 2016). Before the start of the core process of protein purification, researchers should do the following steps to get ready for the process of protein purification. We first need to bring the proteins in a soluble solution to purify them from their cells, tissues, or organs. This can be done by various processes like sonication where sound of high frequency is applied to cells/tissues so that the cell membrane is disrupted and the protein comes out of the cell. After that centrifugation is done where centrifugal force is applied to separate the proteins from cell debris. The heavier particles settle down as a pellet and the supernatant is taken for downstreaming. Following steps are generally used by the researcher for the extraction of proteins.

1.2.1 Protein Precipitation/Salting Out

Due to the availability of hydrophilic amino acids on the surfaces of proteins, they are usually soluble in water because they interact with water molecules. This solubility depends on the pH of the solution and is the function of the ionic strength. Isoelectric point (pI; the point at which there is no net charge, i.e., the charges of their amino acid side groups balance each other) is a unique property associated with proteins. Protein tends to precipitate at their pI, if the ionic strength of a solution is either too lower or too high. The solubility of protein depends on ionic strength. Proteins will precipitate as the ionic strength is increased by adding salt. For this purpose, the most common salt used is ammonium sulfate because it is unusually soluble in cold buffers also (Hyde et al. 2017).

In the salting-out method, salts are added to the extracted protein solution. Proteins may also have hydrophobic amino acids on surface; when present in a solution, hydrophilic amino acids form bonds with the solvent (mostly water) and the hydrophobic amino acids curl inside, but when salt is added it interacts with solvents leaving less solvent for the hydrophilic amino acids, as a result, the proteins precipitate out of the solution by forming aggregates. This process is called salting out. The concentration of salt at which a protein precipitate varies from protein to protein. Therefore, salting out is used to fractionate cellular proteins. For example, fibrinogen (a blood-clotting protein) is precipitated with 0.8 M ammonium sulfate, whereas serum albumin is precipitated with 2.4 M ammonium sulfate. Ammonium sulfate fractionation is the first step in protein purification and commonly used by researchers in the lab because it provides crude protein purification and separates proteins from the non-proteinaceous part. The ammonium sulfate yields a slurry of precipitated protein that is very stable. The protein slurry can be stored in the cold condition for a long period without any detrimental effect on the protein. Ammonia is very reactive with stainless steel; therefore, working with ammonium sulfate in the industry has some problems as steel is the primary building material of industrial purification facilities (Grover and Ryall 2005). Other salts such as sodium sulfate may also be used for the precipitation of proteins, but none of the salts are as good as

ammonium sulfate. Besides salting-out process, proteins can also be forced out of solution with polymers of polyethylene glycol (PEG). PEG is the most frequently used polymer because it is inert and has anti-freezing property. PEG, like ammonium sulfate, tends to stabilize proteins. Increased concentrations of PEG are added with gentle stirring to cold protein solutions. The precipitated proteins can be removed by either filtration or centrifugation. Protein precipitation process typically gives a high fold of the protein and we need 1000-fold or more precipitated proteins before the start of the actual purification process.

1.2.2 Buffer Exchange/Dialysis

Dialysis is also a method of protein purification that has a central role in purification schemes and usually used for achieving high protein amount. The protein must be stored/suspended in a buffer of specific pH/ionic strengths before applying purification techniques. After ammonium sulfate precipitation, proteins are present in a high salt environment. How to remove salt is a massive question because it is difficult to remove salt from the solution? So, we choose selectively permeable membranes (dialysis bags) for the removal of salts. Protein solution is poured into a dialysis bag. Dialysis bag is cellophane tubing that has tiny pores in it through which salt molecules may pass out but protein molecules cannot because of its larger size. When the dialysis bag is filled with a protein solution and placed in a large batch of the buffer, the salt concentration will gradually become equal inside and outside the bag. After some time, the salt concentration inside the bag becomes same as the buffer if the outside buffer is changed often. Now the protein is in the buffer of the choice. Dialysis is a time-taking process that usually takes overnight. This process is effective in separating salts and other small molecules from proteins, while it is not so good in separating proteins from other proteins (Ritchie 2012).

After following these protein preparation steps, the core process of purification starts. There are various methods available for the purification of proteins. In every step of protein purification, there is a loss of product and therefore buffer equilibration steps are very important. On every surface, the protein molecule binds and is inactivated easily because of the shear forces, foaming, or rapid changes in ionic strength. Some methods/process of protein purification is explained briefly here to make the reader aware of the various processes.

1.2.3 Immuno-Affinity Chromatography

This is a technique of choice for beginners until they find out it does not work almost as well as they thought. It depends on the binding affinity of a protein specifically to an immobilized ligand, while the rest of the protein molecules pass through the column. The choice of ligands is antibody and a simple antigen–antibody interaction cause protein purification. The drawback of this method is other proteins may also interact with the antibodies and bind nonspecifically; also antibodies are very

expensive. Immuno-affinity columns are used in this process but usually, this column is only used in late of this process when the volume of the sample has been reduced and most contaminants have considered being removed. The product obtained via this process is an expensive product (Moser and Hage 2010).

1.2.4 Affinity Chromatography

The previously discussed method is based on antigen–antibody interaction, but this method uses other molecules in place of antibody, which is less expensive and very useful. This is known as affinity chromatography where the protein of interest binds to a specific carbohydrate, specific chemicals, or needs a specific cofactor. This method takes advantage of the high affinity of many proteins with their ligands (specific chemical groups). In the method of affinity chromatography, the sample is passed through a matrix or column containing the chemical group to which our desired protein can covalently bond while the rest of the unbounded proteins elute out first, and then to release the protein-bound with the matrix. It is then washed with a buffer and further the desired protein is eluted in almost pure form. We may be capable to complete it by binding them in a column in which carbohydrate (or another cofactor) is immobilized. Protein can then be eluted with a very high concentration of the cofactor or carbohydrate present in a column. Affinity matrices can be used sometimes to mimic the binding sites. There are several examples of useful affinity matrices that helps in the purification of a particular protein, and researcher has to select matrices according to their experimental need (Darcy et al. 2011). This method is a very effective and applicable method of purifying proteins.

1.2.5 Ion-Exchange Chromatography

It is a very appreciated process of protein purification. The same protein can be adsorbed to both cation exchangers (that bind positively charged particles) and anion exchangers (that bind negatively charged particles) by choosing different buffers. This occurs because proteins have different charges at different pHs. Most proteins are negatively charged at physiological pH values (pH 6–8); therefore, most of the proteins are purified on anion exchange columns. At extreme pHs, proteins become inactivated, therefore we should avoid extreme pH conditions. The material of the matrix in the column in ion-exchange chromatography comprises of beads of some inactive material. These inactive materials are often a carbohydrate such as cellulose or dextran. These matrices exist in the form of very fine beads that can be slurred in the buffer and poured into a vertical glass/plastic column formed. Beads adsorb the proteins when the protein solution is pumped onto the top of the chromatography column. The proteins in the column are then eluted according to the firmness with which they attach with the matrix. The proteins that are the most highly charged will bind the tightest with a matrix at that pH. We can either increase the concentration of salt or change the pH to release the proteins in the order of their binding tenacity

(Khan 2012). Both methods (high salt concentration or change in pH) are used in industry, but raising the concentration of salt is the most commonly used method because it is easier to control.

Based on their net charge, proteins are separated in ion-exchange chromatography. If a protein has a net positive charge at pH 7, it will bind to a column of beads usually that have a negative charge on it, whereas a negatively charged protein will not bind to the column bead. Such positively charged bonded protein then can be eluted by increasing the concentration of sodium chloride or another salt in the eluting buffer. This is done because Na^+ competes with positively charged groups on the protein for binding to the column. Proteins that have a low net positive charge density will leave the column first, followed by those proteins that have high charge density. Cationic proteins (positively charged proteins) can be separated on negatively charged CM-cellulose (carboxymethyl-cellulose) columns and, anionic proteins (negatively charged proteins) can be separated by chromatography on positively charged DEAE-cellulose (diethylaminoethyl-cellulose) columns (Acikara 2013).

Using a salt concentration gradient, we elute proteins in an ion-exchange column in research laboratories routinely. When increasing amounts of salt are added to the buffer, there is a continuous steady increase in the ionic concentration going through the column. The proteins then elute/or come off the column matrix when the ionic strength of the buffer neutralizes their charge. The most highly charged molecule comes off last (due to tight binding of the matrix) and the least charged molecules come off first (due to weak binding of the matrix). The most common way to elute columns in the industry is by step elution because this type of gradient is too difficult to control precisely in the industrial setup. The column is carefully rinsed with buffers of increasing ionic strength until the all protein comes off. The same sequence of the procedure is repeated each time with the same amounts of buffer to give reproducible yields and purification of the protein. A single anion exchange column is not enough to purify the protein of interest even with gradient elution; therefore, additional procedures are necessary for final purification of protein.

1.2.6 Size-Exclusion or Gel Filtration Chromatography (SEC/GFC)

It is not an accurate process of specific protein purification as proteins can be of many sizes and the variation in size of protein is used in SEC to separate them. It gives very gross separations is therefore not very effective at the commercial scale. Beads used in this process are made up of a porous matrix into which the proteins get diffused. The pore size of the bead determines the rate at which proteins of various sizes diffuse into it. Because of the low pore size, some proteins (bigger) are completely excluded. We can achieve separation of protein of interest by choosing the correct pore sizes for the protein of interest. The protein samples are poured to the top of a column comprising porous beads (made of an insoluble but highly hydrated polymer such as agarose/dextran/polyacrylamide). Sepharose, sephadex, and bio-gel are the commonly used matrix in this method. These beads are typically 0.1 μm in

Table 1.1 Tabulation of protein purification methods and their respective recovery and level for a fictitious protein

S. no.	Process	Approximate protein remaining after the process (in mg)	Level of protein purification
1.	Tissue homogenate	15,000	1
2.	Salt fractionation	4600	3
3.	Ion-exchange chromatography	1200	9
4.	Size-exclusion chromatography	70	110
5.	Affinity chromatography	2	3000

diameter. Small protein molecules are captured inside these beads, but large molecules cannot. As a result, small protein molecules are dispersed in the aqueous solution both inside the beads and between them, whereas large molecules are situated only in the void volume (solution between the beads). The column (packed with beads) is eluted with buffer. The large protein molecules present in the void volume flow more rapidly through the chromatography column and emerge first during elution. Molecules that are of smaller size take more time and stay last. This procedure requires a long narrow column to achieve protein separation, which is a big obstacle for its use in the industrial processes (Porath and Flodin 1959). If we start protein extraction and purification from tissue homogenate, the approximate amount of protein and purification level in various steps is summarized in Table 1.1.

1.2.7 High-Pressure/Performance Liquid Chromatography (HPLC)

This is an advance method of chromatography. It is highly specific and resolves the proteins minutely. Itis an effective separation platform where the column chromatography method can be used. The column materials are more finely divided in the HPLC method than the above-discussed column materials. The HPLC hence provides an increased surface area for binding of the protein and therefore will give a high-resolution rate of protein purification. High pressure is applied in the columns of HPLC because of the finer materials used to ensure the constant flow rate (Kirkland 1971).

1.2.8 Hydrophobic Interaction Chromatography

Proteins contain both hydrophilic and hydrophobic amino acids. The hydrophobic amino acids are present mostly in the core of the protein structure, i.e., away from the surface, but some hydrophobic amino acids are present on the surface of almost all proteins. These hydrophobic amino acids are used in the column as ligands to separate proteins based on their relative hydrophobicity. The beads used for this

chromatography column are coated with hydrophobic fatty acid chains. The hydrophobic amino acid side groups are not normally exposed to proteins because they attract numerous molecules of water. Therefore, the entire protein is surrounded by the molecules of water, except in the environment of high salt concentration, because in high salt concentration, the hydrophobic areas are exposed and bind to the matrix. The protein is eluted from the column with decreasing concentrations of salt in the buffer. Therefore, we put proteins onto this column in high salt and get them off in a low concentration of salt. This principle makes this method very useful as a next step after proteins are eluted with high salt from ionic exchange columns. The hydrophobic column does not require buffer exchange because the protein is eluted-off stepwise in low ionic strength, so the column is ready for the next purification step without a further buffer exchange (McCue 2009). It saves time as well as the loss of protein.

1.3 Resolving and Display of Protein

Once the protein is purified, it needed to be resolved on the gel to get an idea of quality and approximate molecular weight of protein as this information is very useful for further structure–functions studies. To know the molecular weight of protein, the first thing that comes in our minds is to sequence the amino acids present in the proteins, but it is quite a long, expensive, and tedious task. Therefore, some simple techniques based on electrophoresis have been developed to resolve and visualize the protein. Gel electrophoresis is routinely used to observe the complexity of the protein mixture by separating the proteins on a thin gel of polyacrylamide. The thin gel allows the separation of proteins that unfortunately cannot be scaled up to the production levels without major losses in the efficiency. The basis of protein separation is the same both in electrophoresis and ion-exchange chromatography, i.e., heavier protein moves slowly and lighter protein moves fast on a normal polyacrylamide electrophoresis gel. When the thickness of gel is increased, there is buildup in heat that ultimately causes disturbances in the flow of protein. Therefore, only small protein sample volumes can be applied to the gel electrophoresis. These made electrophoresis impractical for the separations of proteins commercially.

1.3.1 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

It is a very common method in the laboratory and frequently used to analyze proteins during/after protein purification. This method is based on the principle of movement of charged particles in an electric field and used for separating proteins and other macromolecules (DNA and RNA). SDS-PAGE separation is carried out in polyacrylamide gel because the gel serves as a molecular sieve that enhances separation. Charged molecules move in the gel toward the oppositely charged electrode depending upon its charge-to-mass density. Molecules that are small as compared

with the gel pores(sieve) will readily move through the gel, whereas molecules that are much larger than the gel pores are usually immobile. Molecules having intermediate size move through the gel with various degrees of mobility. PAGE is performed in a thin, vertical slab of polyacrylamide, and the direction of movement of the protein molecule is from top to base. Polyacrylamide gels are formed by the polymerization process of bis-acrylamide and acrylamide. Polyacrylamide gel is chosen for electrophoresis because it is readily formed and chemically inert. Proteins are to be in denatured form for running on an electrophoresis gel. PAGE is performed with SDS; therefore, the mixture of proteins is denatured in SDS, and denatured protein has a large net negative charge that is approximately proportional to the protein mass. The negative charge gained because of SDS binding is much greater than the charge on the native protein. When the electrophoresis is complete, the proteins in the gel can be visualized by staining them with Coomassie brilliant blue dye or silver staining, which displays a series of protein bands. If the protein has been labeled radioactively, then the radioactively labeled protein can be detected by placing a sheet of X-ray film over the gel and this procedure is called autoradiography. Large proteins stay at the top and small proteins toward the bottom of the gel. The mobility of most polypeptide chains under SDS-PAGE is linearly proportional to their masses (Schägger 2006).

1.3.2 Two-Dimensional Gel Electrophoresis

Proteins move in SDS-PAGE in the vertical dimension (i.e., in one dimension; from top to bottom), and proteins separated in PAGE are based on their mass only. Proteins can also be separated electrophoretically in two dimensions: first dimensional based on of their pI and second dimension based on their mass. The pI (isoelectric point) is the pH at which a protein has no charge, i.e., the net charge is zero. It means electrophoretic mobility of a protein is zero at pI. The method of protein separation according to their pI is called *isoelectric focusing (IEF).* The pH gradient in the gel (first dimension) is firstly formed by subjecting a mixture of polyampholytes (small multi-charged polymers) having many values of pI (1–14) to the electrophoresis. IEF can readily resolve proteins that differ in pI value by as little as 0.01. It directly means that proteins differing by even a fraction of net charge can be separated. After IEF, proteins are subjected to SDS-PAGE (second dimension) to get very high-resolution separations of protein. IEF combined with SDS-PAGE is known as 2D gel electrophoresis (2DGE).

The sample protein is first subjected to IEF with high voltage in 2DGE. This electrically focused protein gel is then placed horizontally on top of an SDS-polyacrylamide gel slab. Now, the proteins are spread across the top of the polyacrylamide gel according to how far they migrated during IEF. They then again undergo electrophoresis in a perpendicular direction (vertically as per Sect. 3.1) to yield a 2D pattern of the gel in which protein is represented in the form of spots (Fig. 1.2). In 2DGE, proteins are resolved in the horizontal direction (IEF; based on pI) and in the vertical direction (SDS-PAGE; based on mass). Notably, more than a

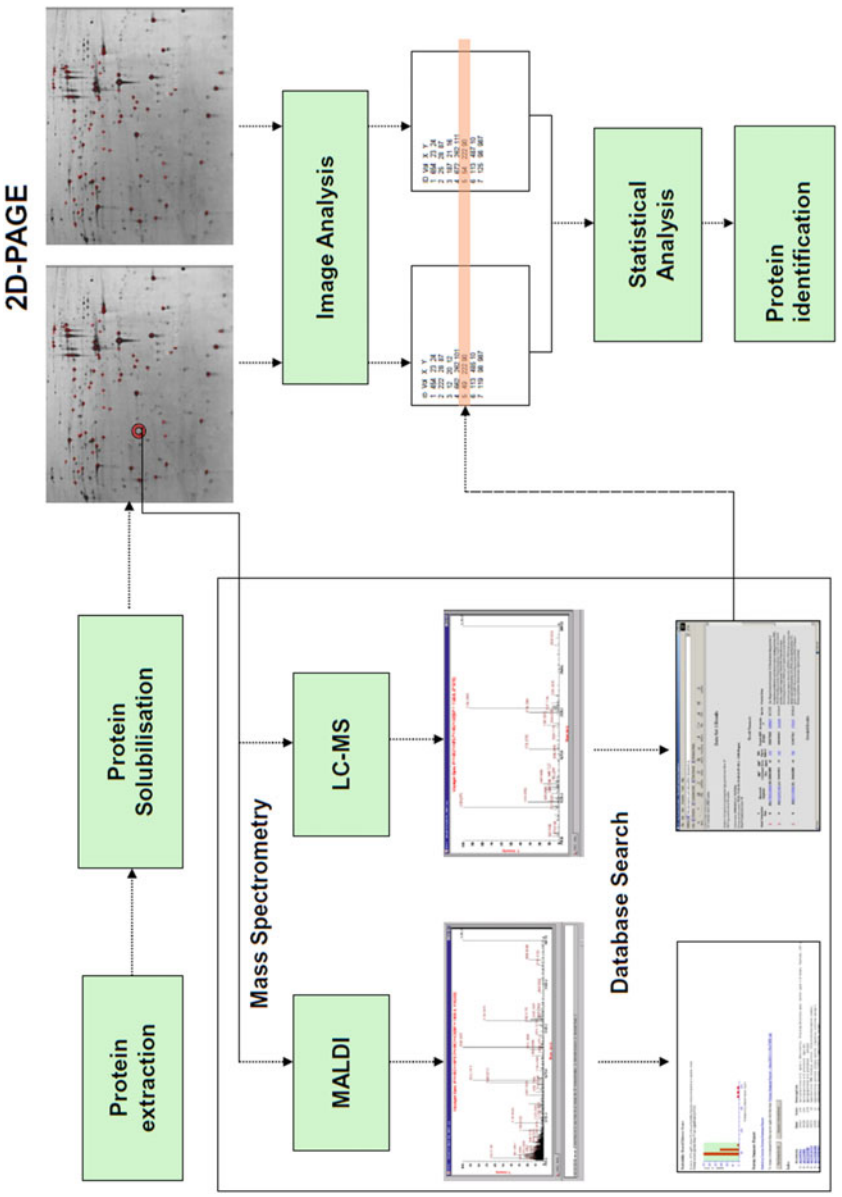


Fig. 1.2 A workflow for proteins isolated from cells for 2D-PAGE followed by mass spectrometry for identification

thousand different proteins can be resolved on the gel in a single 2DGE experiment (Ciborowski and Silberring 2016). Proteins isolated from cells under different physiological conditions are subjected to 2DGE to get them resolved based on their pI and mass. The 2D gel of a protein mixture is followed by an examination by seeing the intensity of the protein spot signals on the gel (Fig. 1.2). In this way, particular proteins can be seen for its increase or decrease in concentration with the response to the physiological state. The 2D gel displays many protein spots that could be further identified by the process of mass spectrometry (MS). It is now possible to identify various cellular proteins at a time by coupling 2DGE with MALDI-TOF (matrix-assisted laser desorption ionization-time of flight: a mass spectrometric technique for protein identification) (Fig. 1.2).

1.4 Methods of Protein Estimation and Assay

After purification, we must estimate the proteins to know its quantity because quantification of protein is an integral part of any laboratory involved in protein purification and other analysis (Fig. 1.1). Protein must be solubilizing in an aqueous solution of buffer before a total protein content estimation. Some precautions are often taken during protein estimation to inhibit microbial growth or to avoid casual contamination of the protein sample by unwanted material such as body oils, dust, hair, and skin. No one reagent and methods can be the ideal or best for the estimation of protein. Each method has its advantages and drawbacks. The choice among available protein estimation to opt is usually based on the compatibility of the protein assay method with the samples. The goodness of the assay method is based on the accuracy, reproducibility, and incubation time desired to complete the assay. Choosing an appropriate assay should be compatible with the protein samples. The assay should be easy to analyze, standard, and associated with good understanding and fewer limitations. Several methods are available for protein estimation, but some methods of choice are described as follows.

1.4.1 Bicinchoninic Acid Method (BCA; Range: 20–200 μg)

It is a calorimetric assay and developed in the year 1985. It is a newer method compared to Lowry assay (the most popular method of protein estimation). The first step in the BCA method is to complex the protein with copper ions. This step is very similar to the Lowry method of protein estimation. The protein bounded with copper chelates further reacts with BCA and gives an intense purple color in the second step. The intensity of the purple color is directly proportional to the amount of protein, and each sample's intensity must be compared with a standard curve drawn in this assay. This assay is very useful for estimation if the protein samples contain detergents $>5\%$. This assay is robustly prejudiced by tyrosine, tryptophan, and cysteine amino acid residues. Chemicals that interact with copper (such as ammonia) may interfere

with the BCA assay because it depends on copper ion coupling for the first reaction (Brown et al. 1998).

1.4.2 Bradford Method (Range: 1–20 μg)

This method is gracefully simple in which negatively charged CBB-250 (Coomassie brilliant blue-250) dye binds with positively charged amino acids present in the protein sample. When the CBB dye is present in solution freely, its absorbance is taken at 465 nm because it is red but when it binds to positively charged basic amino acids of the protein, it absorbs 595 nm light as it turns blue. The absorbance in the protein sample can be then compared to a standard curve. The Bradford method is quick, easy, and stable for up to an hour. However, it can detect proteins larger than 3 kDa. It is very susceptible to detergents like SDS, Triton X-100 unlike the BCA method of protein estimation (Hammond and Kruger 1988).

1.4.3 Folin-Lowry Method (Range: 5–100 μg)

This is a colorimetric-based assay that works in two steps: (a) it made complexes of copper with the nitrogen present in the protein; (b) the complexed tryptophan and tyrosine residue react with FolinCiocalteuphenol reagent. Phosphomolybdotungstic component of this reagent gives an intense blue-green color which absorbs light at 650–750 nm. Unlike BCA, the Folin-Lowry method is an endpoint assay with a stable result. It means that we can estimate the amount of protein with one assay by comparing it with a previous protein standard curve. Unfortunately, this assay is not well-suited with lots of common chemicals like β-mercaptoethanol, DTT, reducing agents, EDTA, Tris, carbohydrates, and potassium/magnesium ions (Lowry et al. 1951).

1.4.4 Kjeldahl Method (Range: 1–100 μg)

It is an ancient (>130-year-old) method where nitrogen is measured in the protein sample. In the Kjeldahl method, nitrogen is converted into ammonia through a series of frightening steps involved in steam distillation, heated sulfuric acid, and back-titration with sodium hydroxide. After following all these steps, the purified nitrogen of protein sample is weigh out. Further, it is assumed that the original protein sample was 16% nitrogen, then back-calculate the total amount of protein. It is a monotonous and time-consuming method of protein estimation. The Kjeldahl method needs at least 1 g of protein sample to estimate the amount of protein present in a sample. One gram of protein is a huge amount that makes this method highly unreasonable and impractical for the molecular biologists (Sáez-Plaza et al. 2013).

1.4.5 Ultraviolet Absorption Method (Range: 20–3000 µg)

This process estimates the protein amount by determining the characteristic absorption of tryptophan and tyrosine at 280 nm. This is a very simple method but not very much reliable because every protein has a different amount of tryptophan and tyrosine amino acids. Several other molecules (alcohols, certain buffer ions, and nucleic acids) also interfere with the result because all absorb at 280 nm. The result obtained by this method may be nonspecific if the nucleic acid is available in the protein sample (Noble 2014).

It is important to understand the basic idea of different protein estimation methods. A protein estimation method tells us how much protein is present in a sample, but it does not tell how much protein is available unless it is pure. We have to have a prior idea about how much a protein is available in a sample and how much contaminating material is also present along with protein. Therefore, the single and most important idea central to protein estimation/purification is a specific activity. All assays (described above) are only estimating the amount of protein present. A useful assay or method of protein estimation can be very simple or complex, but it should be relatively rapid and reproducible with good precision.

1.5 Storage of Proteins

Proteins are typically unstable if they are not stored in their native environments after purification. Cellular compartments and extracellular fluids provide native environment to the proteins. Each protein needs a specific environment and requirements once it is taken out from its normal biological niche. Protein can rapidly lose its ability to perform specific functions if it does not satisfy these requirements. Proteins may lose its activity because of proteolysis and aggregation. If it is stored in less than optimal storage conditions, its activity must be lost. Therefore, there is a regular need to store purified proteins so that its original structural integrity and activity could be retained for an extended period of time. The shelf life of a protein varies from a few days to more than a year. It totally depends on the nature of the protein and the storage conditions followed after purification. Protein solution may be stored at 4 °C or at −20 °C (25–50% glycerol or ethylene glycol) or in the deep-frozen form (−80 °C to liquid nitrogen temperature) or in lyophilized form. Typically shelf life of protein is 1 month in 4 °C, 1 year in glycerol or ethylene glycol, and many years in deep-frozen form. The addition of an antibacterial agent is required while storing at 4 °C, but nothing is required to add in protein solution while storing it in frozen condition (Arora 2013). Many times, we can use the samples if they are stored at 4 °C or in 25–50% glycerol/ethylene glycol. But deep-frozen protein can be used once. Repeated freeze-thaw cycles are avoided as regular freeze-thawing decrease the stability of the protein. Proteins are stored in autoclaved container, or polypropylene tubes. Less than 1 mg/mL protein stock is more prone to inactivation or loss of activity. Compounds (Table 1.2) used in storage may cause protein solutions to lengthen the shelf life.

Table 1.2 Compounds used for protein storage and their role

S. no.	Compound	Role	Concentration used
1.	Glycerol or ethylene glycol	Cryoprotectants	25–50%
2.	Protease inhibitors	Prevent proteolytic cleavage	In mM concentration
3.	Sodium azide	Antimicrobial substance	0.02–0.05% (w/v)
4.	Thimerosal	Antimicrobial substance	0.01% (w/v)
5.	EDTA	Metalloprotease that make protein more stable at colder temperatures	In mM concentration
6.	Dithiothreitol (DTT) and 2-mercaptoethanol (β-BME)	Reducing agent	In mM concentration

1.6 Influence of Additives on Protein During Storage

1.6.1 Presence of Bacteria/Microbes

If a researcher does not work in a sterile environment, the entire protein sample may get contaminated with various microbes present in the environment. An antibacterial agent such as sodium azide (NaN_3) should be added in the sample in a working concentration of 0.02–0.05% (w/v), even if protein solutions are kept at 4 °C (Hippensteel et al. 2016).

1.6.2 Formaldehyde/Glutaraldehyde

This reagent is used for cross-linking of proteins; therefore, precaution is needed while using this reagent in purification. They are highly vulnerable to the nucleophilic attack of the amino groups present in the protein. A consequent covalent bonding and cross-linking could occur because of this interaction (Migneault et al. 2004).

1.6.3 Cycles of Freeze-Thaw

Repeated cycles of freeze-thaw usually degrade proteins. As a rule of thumb, there is no way to freeze the same batch of protein twice. Prepare various aliquots of the same protein samples for the safer side for storage and use once. Do not store/freeze the protein sample repeatedly (Ji et al. 2017).

1.6.4 Organic Solvents and pH

Organic solvents mostly denature the protein; therefore, loss of function takes place. But it does not affect the primary structure of the protein. Every protein shows its biological activity or function at an optimal pH. Even slight changes in the pH affect the protein activity seriously. A change of pH from 7.2 to 6.2 means >90% loss of protein activity. A strong basic or acidic pH denatures the proteins rapidly. The physiological pH of most of the proteins is 7.2–7.4 (Talley and Alexov 2010).

1.6.5 Proteases/Peptidases

Proteases and peptidases are enzymes that degrade the proteins and peptides. Most of these enzymes have optimum activity at a temperature of 37 °C approximately. At this temperature, proteins are degraded by the enzyme proteases within few minutes. As we know that proteases show reduced activity at lower temperatures, but it does not mean that proteases are not active at 4 °C temperature. Protease inhibitors must be added to the protein solution if we go for long-term storage of protein. Proteases are found not only on the human skin but also in the air. Therefore, a protein sample can be easily contaminated during handling. Proteases are produced by bacteria, and infected samples are prone to protease contamination (Razzaq et al. 2019).

1.6.6 Protein Concentration

During storage, protein may bind to the material of the vessel (microfuge tube). When low protein concentrations (<1 mg/mL) are kept in the vessel, some protein may bind with storage vessel material, resulting in the loss of protein in the sample (Weikart et al. 2017).

1.6.7 Reducing Agents

β-Mercaptoethanol (BME), dithiothreitol (DTT), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), etc., are strong reducing agents. They all break the disulfide bonds present in proteins and cause loss of the tertiary/quaternary structure of proteins. They are most commonly used in SDS-PAGE for sample preparation. Such reducing agents denature every protein that contain disulfide bonds (Ritchie 2012).

1.6.8 Salt Conditions and Urea

Proteins are present in a physiological environment with salts. Normally they are present at a weak balance of salt concentrations. Too low or too high salt

concentrations may precipitate the proteins. PBS (phosphate buffer saline: equivalent to physiological buffer) is commonly used to dissolve protein. Pure water is not recommended to dissolve protein. At high concentrations (> 6 M), urea denatures proteins. Like urea, guanidinium hydrochloride and lithium perchlorate are other chaotropic agents (that disrupt the hydrogen bonding network between water molecules) that devastate the secondary and tertiary structure of proteins by influencing hydrogen bonding, hydrophobic effects, and others (Tripathi 2013; Zhang et al. 2017).

1.6.9 Temperature

The optimum temperature of mammalian proteins is around 37 °C, and they show their biological function accurately at this temperature. Most of the mammalian proteins denature above 43 °C temperatures. The complete denaturation of protein takes place within 1–2 h at 55 °C and even 2 min can completely denature protein at 95 °C. However, proteins also denature at room temperature, but this denaturation is mainly because of the involvement of other factors. In their cellular environment, a protein is protected by other proteins (chaperones), but chaperons are not present in the solution of purified protein. Therefore, the protections of protein must be ensured while storing and working (Vieille and Zeikus 2001).

1.7 Factors Affecting the Stability of Proteins

Easily proteins lose their biological function and overall stability if native environment is not provided. There are the number of functional requirements associated with proteins. Hydration explains the solubility, dispersibility, gelation, and viscosity of a protein molecule. Aggregation, gelation, viscosity, and extrudability are the rheological features of the protein. Factors that affect viscosity are the concentration of protein, ionic strength, pH, and temperature. Stabilization of protein is depending on viscosity, solubility, rate of diffusion, protein flexibility, net charge, and protein hydrophobicity (Cheung and Mehta 2015). Flavor, odor, texture, and color are the sensory properties of a protein. To ensure these properties (Table 1.3) of the protein,

Table 1.3 Functional requirements of protein ingredients

S. no.	Property	Functional attributes of proteins
1.	Hydration	Solubility, dispersibility gelation, viscosity
2.	Rheological	Aggregation, gelation, viscosity extrudability
3.	Sensory	Flavor, odor, texture, color
4.	Surfactant	Baking, emulsion, foaming, whipping
5.	Textural	Adhesion, aggregation, gelation, viscosity
6.	Visual	Opacity, turbidity, color
7.	Others	Comparable with processing conditions and other ingredients

its storage is an important issue, otherwise, its stability gets affected. A protein establishes functionality by interacting with other macromolecules. These interactions may comprise solute molecules, solvent molecules, substances that are dispersed in the solvent, and other protein molecules. The energies and forces involved in the maintenance and achievement of the native structure of protein must be selected to depict the forces involved in these interactions. Protein must interact with other components of the systems, and it may affect the functionality of the protein. Sometimes the presence of other molecules in a protein solution allows the interaction. But interactions commonly require an input of energy into the surroundings, system, and interacting molecules. A complete discussion on the forces involved in these interactions is beyond the scope of this chapter, therefore, the discussion is limited to the factors affecting the stability of proteins. Proteins exist in the lowest kinetically achievable free energy state. The structure and conformation of the protein highly depend on the environment if the environmental conditions change, the configuration and conformation of protein will change. Environmental factors include temperature, pH, ionic strength, dielectric constant, and other molecules. Lowering of free energy of protein involves the removal of hydrophobic groups from the aqueous environment that can have large influences on the structure of the protein (Abriata et al. 2016).

Protein molecules contain cross-linking of a covalent (disulfide bonds) and non-covalent (salt bridges) interactions. These cross-linking lowers the conformational entropy of the molecule, which is compensated by lowering in the binding energy. The occurrence of cross-linking adds stability greatly to the structure of a native protein and makes the protein molecules resistant to denaturation/unfolding. One of the driving forces of denaturation is an increase in conformational entropy. When the denaturing agents are removed, the gain in conformational entropy is a large driving force to maintain the denatured state (Tripathi 2013). The increase in entropy is much less and thus in contrast to a highly cross-linked protein, we cannot assume the same degree of random conformations. This statement is the answer to the question of why does a protein is often resistant to denaturation that contains a numerous disulfide bonds (Hammann and Schmid 2014). The structures achieved by the proteins are very dynamic rather than rigid. There is rotational freedom about many of the bonds within the molecule of protein, and the total free energy of the native protein structure gets lowered due to entropy gain of this freedom. Because of changes in storage condition, a protein may adopt a variety of conformations. These new structures of the protein differ slightly from the native conformation and always lead to a situation where there is an increase in the free energy of the system. Thus, the native structure of a protein has the lowest free energy. Slight changes in the environment shifted a protein toward alternate structures (lowest free energy) and thus lead to denaturation of the protein (Yang et al. 2013).

1.8 Effect on the Dynamics of Protein

Protein dynamics are the different conformational states that a protein adopts. Protein form unique structures that are determined by their amino acid sequences. However, protein is not a static object, but it exists in a group of conformations (states) depending on their environment. Transitions between these states of protein occur on a variety of time scales (nanoseconds to seconds), length scales (Å to nm), and linked to functionally relevant. Detailed knowledge of dynamics is required to understand the function of the protein. Despite substantial experimental progress, it is difficult to know about the dynamics of protein (Sikosek and Chan 2014). Molecular dynamics (MD: a simulation technique) offer the only routine means to obtain dynamic information at the level of atom on the timescales of nanoseconds to microseconds (ns to ms). Sampling techniques beyond MD are necessary to enhance the conformational sampling of large proteins and their assemblies, even with the current development of computational power. High throughput experimental techniques like X-ray crystallography, nuclear magnetic resonance (NMR), and electron microscopy have made considerable progress in deciphering the 3D structure of proteins. Protein function is a dynamic process that involves the conformational transitions and structural rearrangements between the stable structure. Such dynamic processes are very hard to study at the experimental level. No doubt, proteins perform a remarkably large number of functions in the body and control/regulate various cellular phenomena. The study of the dynamic behavior of proteins opens a new arena to understand functional protein conformation where more than one stable structure is found. It has become more and more obvious that many proteins function by conformational changes between stable structures (Klepeis et al. 2006).

The long-term stability of protein affects the structure–function-–dynamics of protein. Upon lyophilization, proteins may unfold reversibly/irreversibly and acquire conformations susceptible to the degradation of protein during the storage. The process of lyophilization generates various stresses, including ice crystal formation, increased local solute concentration, and pH changes that can damage protein structure. Stabilizing agents (e.g., sucrose and trehalose) are often added to protect protein structure during drying or freezing. But proteins can exhibit a reversible or irreversible change in structural properties, which include a change in secondary structure, tertiary structure, conformational dynamics, and/or conformational changes upon lyophilization (Moorthy et al. 2015). There is some experimental method that shows considerable promise for mapping local protein–protein, protein–water, and protein–excipient interactions in lyophilized solids with high resolution. The protein binds with water and the factors that affect the water-binding capacity of proteins are amino acid composition, ion species, ionic concentration, protein conformation, surface polarity/hydrophobicity, pH, and temperature. If a protein sample is put into liquid nitrogen directly (flash freezing), protein gets precipitated and activity is lost during rapid freezing. But even in this situation, disulfide bond and secondary structures would be fine. Several techniques are available for the study of protein dynamics. FTIR, DSC, NIR, DRS, solid-state NMR, neutron

scattering, and Raman spectroscopy could provide complementary information about protein conformations and structural changes (Alberts et al. 2002). Despite the development of advanced techniques, our experimental knowledge of protein dynamics is still limited.

1.9 Conclusions

Proteins are essential components of biological processes and highly complex biomolecules with specific functions. This complexity explains why its understanding is so important. In this chapter, an overview of its extraction, purification, estimation, storage, and dynamics is discussed. Knowing these facts related to protein is very essential because biological function, molecular stability, and dynamics of a protein depend on a variety of factors. If someone plans to use proteins in their work, the individual should be highly aware of the factors discussed in this chapter and work consequently. Therefore, this book chapter is written to compile knowledge on protein extraction, purification, estimation, storage, and effect on its dynamics.

Acknowledgment The author is thankful to the National Institute of Technology (NIT), Raipur (CG), India.

References

Abriata LA, Spiga E, Peraro MD (2016) Molecular effects of concentrated solutes on protein hydration, dynamics, and electrostatics. Biophys J 111(4):743–755

Acikara OB (2013) Ion-exchange chromatography and its applications. IntechOpen. https://doi.org/10.5772/55744

Ahuja M (2006) Life sciences. (2 Vols.). Gyan Publishing House, New Delhi

Alberts B, Johnson A, Lewis J et al (2002) The shape and structure of proteins. In: Molecular biology of the cell, 4th edn. Garland Science, New York

Arentson-Lantz E, Clairmont S, Paddon-Jones D et al (2015) Protein: a nutrient in focus. Appl Physiol Nutr Metab 40(8):755–761

Arora M (2013) Cell culture media: a review. Labome Mater Methods 3:175

Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry, 5th edn. W H Freeman, New York. Section 4.1—The purification of proteins is an essential first step in understanding their function

Brown RE, Jarvis KL, Hyland KJ (1998) Protein measurement using bicinchoninic acid: elimination of interfering substances. Anal Biochem 180(1):136–139

Cheung PCK, Mehta BM (2015) Handbook of food chemistry. Springer-Verlag, Berlin Heidelberg

Ciborowski P, Silberring J (2016) Proteomic profiling and analytical chemistry, 2nd edn. Elsevier, Amsterdam, pp 1–298

Coskun O (2016) Separation techniques: chromatography. North Clin Istanb 3(2):156–160

Darcy E, Leonard P, Fitzgerald J, Danaher M, O'Kennedy R (2011) Purification of antibodies using affinity chromatography. Methods Mol Biol 681:369–382

Gopal GJ, Kumar A (2013) Strategies for the production of recombinant protein in *Escherichia coli*. Protein J 32(6):419–425

Grover PK, Ryall RL (2005) Critical appraisal of salting-out and its implications for chemical and biological sciences. Chem Rev 105:1–10

Hammann F, Schmid M (2014) Determination and quantification of molecular interactions in protein films: a review. Materials (Basel) 7(12):7975–7996
Hammond JB, Kruger NJ (1988) The Bradford method for protein quantitation. Methods Mol Biol 3:25–32
Hippensteel E, Stumpf E, Langhorn J (2016) Effect of sodium azide concentration on wear and bacteria growth in a hip simulation test. Front Bioeng Biotechnol. Conference Abstract: 10th World Biomaterials Congress. https://doi.org/10.3389/conf.FBIOE.2016.01.02244
Hyde AM, Zultanski SL, Waldman JH, Zhong Y, Shevlin M, Peng F (2017) General principles and strategies for salting-out informed by the Hofmeister series. Org Process Res Dev 21 (9):1355–1370
Institute of Medicine (2005) Dietary reference intakes for energy, carbohydrate, Fiber, fat, fatty acids, cholesterol, protein, and amino acids. The National Academies Press, Washington, DC. https://doi.org/10.17226/10490
Ji X, Wang M, Li L, Chen F, Zhang Y, Li Q, Zhou J (2017) The impact of repeated freeze-thaw cycles on the quality of biomolecules in four different tissues. Biopreserv Biobank 15 (5):475–483
Khan HU (2012) The role of ion exchange chromatography in purification and characterization of molecules. IntechOpen. https://doi.org/10.5772/52537
Kirkland JJ (1971) High speed liquid-partition chromatography with chemically bonded organic stationary phases. J Chromatogr Sci 9:206–214
Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE (2006) Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 19(2):120–127
Lodish H, Berk A, Zipursky SL et al (2000) The three roles of RNA in protein synthesis.molecular cell biology, 4th edn. W.H. Freeman, New York. Section 4.4
Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193(1):265–275
McCue JT (2009) Theory and use of hydrophobic interaction chromatography in protein purification applications. Methods Enzymol 463:405–414
Migneault I, Dartiguenave C, Bertrand MJ, Waldron KC (2004) Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. BioTechniques 37:790–802
Moorthy BS, Iyer LK, Topp EM (2015) Characterizing protein structure, dynamics and conformation in lyophilized solids. Curr Pharm Des 21(40):5845–5853
Moser AC, Hage DS (2010) Immunoaffinity chromatography: an introduction to applications and recent developments. Bioanalysis 2(4):769–790
Noble JE (2014) Quantification of protein concentration using UV absorbance and Coomassie dyes. Methods Enzymol 536:17–26
Porath J, Flodin P (1959) Gel filtration: a method for desalting and group separation. Nature 183:1657–1659
Razzaq A, Shamsi S, Ali A et al (2019) Microbial proteases applications. Front Bioeng Biotechnol 7:110
Ritchie C (2012) Protein purification. Labome Mater Methods 2:134
Sáez-Plaza P, Navas MJ, Wybraniec S, Michałowski T, Asuero AG (2013) An overview of the Kjeldahl method of nitrogen determination. Part II. Sample preparation, working scale, instrumental finish, and quality control. Crit Rev Anal Chem 43:224–272
Schägger H (2006) Tricine–SDS-PAGE. Nat Protoc 1(1):16–22
Sikosek T, Chan HS (2014) Biophysics of protein evolution and evolutionary protein biophysics. J R Soc Interface 11(100):20140419
Talley K, Alexov E (2010) On the pH-optimum of activity and stability of proteins. Proteins 78 (12):2699–2706
Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91

Urry LA, Cain ML, Wasserman SA, Minorsky PV, Reece JB (2017) Campbell biology, 11th edn. Pearson, San Francisco, CA

US Food and Drug Administration (1991) Biotechnology Inspection Guide (11/91). https://www.fda.gov/biotechnology-inspection-guide-1191. Accessed 24 Feb 2020

Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev 65(1):1–43

Weikart CM, Klibanov AM, Breeland AP, Taha AH, Maurer BR, Martin SP (2017) Plasma-treated microplates with enhanced protein recoveries and minimized extractables. SLAS Technol 22 (1):98–105

Yang LQ, Ji XL, Liu SQ (2013) The free energy landscape of protein folding and dynamics: a global view. J Biomol Struct Dyn 31(9):982–992

Zhang C, Yanga M, Zhao K (2017) Insight into the effect mechanism of urea-induced protein denaturation by dielectric spectroscopy. Phys Chem Chem Phys 47:1–9

Experimental and Computational Methods to Determine Protein Structure and Stability

2

Nachiappan Mutharasappan, Guru Ravi Rao, Richard Mariadasse, Saritha Poopandi, Amala Mathimaran, Prabhu Dhamodharan, Rajamanikandan Sundarraj, Chitra Jeyaraj Pandian, and Jeyakanthan Jeyaraman

Abstract

Proteins are versatile biological macromolecules that are involved in many essential processes and basic functions of a cell, including catalytic activity, storage, transport, cell structure, metabolism, cell signaling, and immunity. The functions of proteins are dictated by their structures. For instance, the shape, catalytic activity, and specificity of enzymes depend on both the sequence of amino acids in their active site to which the substrate or drug binds and the nature of protein folding. The stability of protein will determine if a protein is in native folded conformation or the unfolded or denatured state. The key role of drug designing is to enhance protein stability since the marginal stability of a protein could cause loss of protein function, increased degradation, and difficulty in synthesizing protein-based drugs. The folded structure of a protein is stabilized by several atomic interactions such as electrostatic, hydrophobic, van der Waals, disulfide, and hydrogen bonds, while the entropic or non-entropic interactions dominate the unfolded protein conformations. This chapter provides an overview of the techniques to determine the structure and stability of proteins addressing the principles involved in structure prediction with specific highlights on widely used experimental methods and computational techniques, namely protein purification techniques, biophysical/biochemical characterization of proteins, protein structure determining methods, factors contributing to protein stability, and conformational analysis of protein folding. This combination of advanced experimental and computational approaches in predicting the protein structure and

N. Mutharasappan · G. Ravi Rao · R. Mariadasse · S. Poopandi · A. Mathimaran · P. Dhamodharan · R. Sundarraj · J. Jeyaraman (✉)
Structural Biology and Bio-Computing Lab, Department of Bioinformatics, Science Block, Alagappa University, Karaikudi, Tamil Nadu, India

C. Jeyaraj Pandian
Department of Biotechnology, Dr. Umayal Ramanathan College for Women, Alagappa University, Karaikudi, Tamil Nadu, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_2

measuring its stability serves to an exciting future in drug designing and stability engineering.

Keywords

Protein purification · Chromatography · Structure determination · Protein characterization · Protein stability · Protein folding

2.1 Protein Purification Techniques

Protein purification is a mandatory step to study the structure, function, and interactions of proteins. Protein purification methods should not affect the properties of proteins, especially the sequence and size of the component polypeptides. Protein purification processes separate the desired proteins from all the other proteins and non-proteins of the mixture. Proteins can be purified both by analytical and preparative methods. Analytical methods detect and identify the desired protein in a mixture, while the preparative methods produce large amounts of proteins for structural biology study and industrial applications. Proteins that make up less than 0.1% of the dry weight of any tissue could be brought to 98% purity by protein purification techniques. Purification steps vary for each protein, and the appropriate purification methods have to be evaluated by different accessible techniques until the developed procedure could reproducibly yield highly purified biologically active protein.

2.1.1 General Aspects of Protein Purification

Protein purification steps exploit differences in the solubility, size, net charge, polarity, and binding specificities of proteins (Table 2.1).

Most of the protein purification techniques are performed from 0 to 4 °C to eliminate temperature-dependent processes such as unfolding and denaturation of the protein. Most of the protein sources are whole cells in which the desired proteins

Table 2.1 Protein purification methods

S. no	Characteristics of proteins	Purification method
1.	Solubility	Salting out
2.	Size	Ultracentrifugation
		Gel filtration chromatography
		SDS-PAGE
3.	Charge	Ion-exchange chromatography
		Isoelectric focusing
		Electrophoresis
4.	Polarity	Hydrophobic interaction chromatography
5.	Binding property	Affinity chromatography

could be less than 0.1% of dry weight. The molecular cloning methods facilitate the expression of genetically altered or desired protein to be expressed at high levels in a host microorganism, and the recombinant proteins could constitute approximately 40% of the total cell protein of that microorganism.

The first step in protein purification is the isolation of protein outside the cell into a solution. Most cells require mechanical disruption, cell lysis by crushing or grinding to release their contents, which is followed by centrifugation or filtration to remove the insoluble particles. The target proteins may be a cytoplasmic protein or a membrane-bound protein. The membrane-bound target proteins can be recovered by solubilizing the membrane proteins using detergents or organic solvents. The proteins that are purified by various methods based on solubility, size, charge, polarity, and binding specificity must be stabilized to prevent irreversible damage when exposed to experimental conditions.

2.1.2 Stabilizing Proteins

The following factors should be controlled in all the stages of protein purification to enhance the optimal stability of the purified protein.

(i) **Temperature:** Each protein varies in thermal stability. Few proteins denature at lower temperatures while most of the proteins denature at a higher temperature. Therefore, the standard protein purification process is performed at a temperature from 0 to 4 °C.
(ii) **pH:** Buffer solution at a specific pH range is used for protein purification to prevent structural disruption.
(iii) **Surface adsorption:** Protein structures are used in relatively concentrated form to minimize foaming and to prevent its denaturation during contact with surfaces such as plastics, glass, or air–water interface.
(iv) **Degradative enzymes:** Degradative enzymes include proteases and nucleases that are released during the cell lysis steps of the protein purification process. These enzymes have to be inhibited or inactivated by adjusting the pH and temperature or by adding a compound that blocks their activity specifically. But this change in pH or temperature should not have an adverse effect on the target proteins.
(v) **Storage:** To prevent microbial contamination and oxidation of proteins during storage, the protein solutions are generally stored under argon gas or nitrogen or frozen at −196 °C (or) −80 °C.

2.1.3 Quantification of Proteins

Purification of target proteins requires quantitative detection of protein in a solution using quantitative assays specific for the target protein. The most widely used protein quantification assays include the rate of formation of product (proportional to the

amount of enzyme), coupled enzymatic reactions, radioimmunoassay (RIA), enzyme-linked immune-sorbent assay (ELISA), absorbance spectroscopy, and Bradford assay.

2.1.4 Purification of Proteins Based on Solubility

Due to the presence of multiple charged groups in proteins, the protein solubility depends on the concentration of the dissolved salts, pH, temperature, and polarity of the solvent. Salt is added in the "salting in" process in which the solubility of protein is increased at lower ion concentration. The multiple ionic charges of protein are shielded by additional ions, which are weakened by attractive forces among protein molecules. The addition of sulfate salts further decreases the solubility of proteins, and the phenomenon is called salting out. Salting out occurs due to competition between salt ions and other dissolved solutes. At higher salt concentration, added ions are solvated, and therefore, only less amount of bulk solvent is available to dissolve proteins. Salting out is the basis of most widely used protein purification methods as different proteins precipitate with varying concentrations of salt. Therefore, the unwanted protein has to be removed from the solution by adjusting the concentration of salt in a solution containing a protein mixture to just below the precipitation point of a specific protein. Further, after eliminating the precipitated protein by centrifugation or filtration, the concentration of salt in the solution is increased to precipitate the target protein. This method results in the production of purified proteins in large quantities. Ammonium sulfate is widely used for salting out of proteins due to its high solubility in water (up to 3.9 M in water), which favors the preparation of solution with high ionic strength. The pH of the process should be adjusted to near the isoelectric point of the target protein as the proteins are least soluble at their isoelectric point where the net charge is zero.

Example:

Protein	pI
Insulin (Bovine)	5.4
Hemoglobin (Human)	7.1
Serum albumin (Human)	4.9
Ribonuclease A (Bovine)	9.4
Lysozyme (Hen)	11.0

2.1.5 Purification of Proteins Based on Size

Purification methods of protein based on size include ultrafiltration, gel filtration chromatography, and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (Nachiappan et al. 2018).

2.1.5.1 Ultrafiltration

The rate at which the protein sediments is by ultrafiltration technique which is based on the shape of the protein and density of the solution. Proteins are separated based on mass by ultracentrifugation. The analytical ultracentrifuge is used for characterizing non-covalently associated subunits of proteins. In preparative ultracentrifugation, the density of solution increase from the top to bottom of the centrifuge tube. In zonal ultracentrifugation, the protein solution is layered on the top of the preformed sucrose gradient, and protein fractions are collected as per their sedimentation coefficient. In equilibrium density centrifugation, the protein solution is dissolved in concentrated cesium chloride that forms a density gradient during high spin and form bands of protein components based on their densities. The centrifuge tube is punctured to collect the fractions of separated proteins (Laue and Stafford 1999).

2.1.5.2 Gel Filtration or Molecular Sieve or Size Exclusion Chromatography

Protein molecules are separated based on their size and shape by gel filtration chromatography. Gel beads consisting of gel mixture and enclosing an internal solvent space are used as the stationary phase. The size of the pores is determined by the cross-linking efficiency of the polymers of the gel. When an aqueous solution of the proteins of different sizes is passed through the column (molecular sieve), the molecules larger than the pores traverse the column rapidly when compared to the small molecules. The larger molecules are eluted first and collected as fractions, and higher volumes of solvent are needed to elute small molecules. The elution position and molecular mass of known protein aid in identifying the mass of unknown proteins.

2.1.5.3 SDS-PAGE

SDS-PAGE is an electrophoretic technique used to purify and analyze proteins. Sodium dodecyl sulfate (SDS) is a detergent that interferes with the hydrophobic interactions in protein, which usually stabilizes the protein structure. SDS denatures proteins, masks intrinsic charge of proteins and proteins when treated with SDS display rod-like structure. This results in a similar charge-to-mass ratio and shape of all SDS-treated proteins. Therefore, SDS separates proteins based on gel filtration and gel molecular mass. The relative mobility of protein varies linearly with the logarithms of molecular mass; using molecular markers the molecular mass can be predicted with ~10% accuracy (Janson 2012). Molecular masses of the protein subunits can also be determined by SDS-PAGE as SDS disrupts the non-covalent interactions among polypeptides. The link between the subunits in a polypeptide formed by disulfide bonds can be determined using SDS-PAGE via preparing protein samples both in the absence and presence of the reducing agent 2-mercaptoethanol which breaks the disulfide bond.

2.1.6 Purification of Proteins Based on Charge

Proteins can be separated based on charge by ion-exchange chromatography, isoelectric focusing, and polyacrylamide gel electrophoresis.

2.1.6.1 Ion-Exchange Chromatography

Proteins have both negative and positive charges that can bind with both anion and cation exchangers, respectively. The frequently used anion and cation exchangers for protein purification are diethylamine ethyl (DEAE) groups and carboxymethyl (CM) groups, respectively. Agarose-based and cellulose-based resin are the most widely used matrix materials in the ion-exchange chromatography for protein purification. The binding affinity of target protein in ion-exchange chromatography depends on the presence of other ions which would compete with protein to bind with the ion exchangers and also depend on the pH of the solution that influences the charge of the target protein. The desired proteins to be separated are applied to the column after dissolving in buffer with suitable salt concentrations and pH. Then the column is washed with buffer, and the proteins with higher affinities of ion exchanger move slowly through the column when compared to the protein that binds with lower affinities. The effluent of the column is collected in a series of fractions, and the bound proteins are eluted using eluent with higher pH or concentration of salt, which reduces the affinity of the matrix that binds to the protein. The obtained effluents are monitored at an absorbance of 280 nm for the presence of proteins, which can further be tested with a more specific assay for the target protein.

2.1.6.2 Isoelectric Focusing

Charged groups of both polarities are present in proteins, and therefore there is a specific isoelectric point (pI) for each protein. At an isoelectric point, the net charge of the protein is zero, and so it is immobile in the electric field. Isoelectric focusing (IEF) is an electrophoretic technique in which each protein is focused as bands at their respective pI during electrophoresis with a stable pH gradient. Two-dimensional gel electrophoresis can be developed by combing IEF and SDS-PAGE in which protein is separated by IEF in one direction and further separated by SDS by electrophoreses in perpendicular direction. This method results in an develops array of spots representing each protein, and up to 5000 proteins can be separated in a single 2D gel electrophoresis (Roe 2001). All of the expressed proteins of a cell can be cataloged by 2D gel electrophoresis with aparticular focus on the location, quantification, interaction, modification, and activities of proteins. Spots of target proteins after staining with appropriate stain can be visualized, excised, destained, and eluted from the gel and used for characterization by mass spectrometry. Several reference 2D gels are available in the web-accessible databases for many organisms and tissues that can be used to compare, identify, and detect the amount of component proteins in the sample.

2.1.6.3 Polyacrylamide Gel Electrophoresis

Proteins are separated based on electric charge and shape or size by agarose gel or polyacrylamide gel electrophoresis, respectively. Electrophoresis is different from the gel filtration process because the electrophoretic mobility of larger molecules is lesser than the mobility of small molecules with the same charge density. At high pH, the proteins have a net negative charge and move toward the anode in gel electrophoresis. Proteins of similar size migrate as a single band in the gel electrophoresis, which can be visualized by staining techniques such as Coomassie brilliant blue staining, autoradiography, immuno-blotting, and Western blotting, etc.,. Based on the visualization technique and dimensions used, the samples containing less than one ng of protein can be separated by gel electrophoresis (Tanford and Reynolds 2001).

2.1.7 Purification of Proteins Based on Polarity (Hydrophobic Interaction Chromatography)

This chromatography is based on hydrophobic interactions between octyl and phenyl groups in the chromatographic matrix and proteins in the mixture. Non-polar groups on the protein surface interact with the hydrophobic groups at high salt concentration, and both groups are excluded using a polar solvent. Eluent is an aqueous buffer with increasing detergent concentration (disrupt hydrophobic interaction), decreasing salt concentration, and changing the pH.

2.1.8 Purification of Proteins Based on the Binding Property (Affinity Chromatography)

Based on the binding property, the proteins are purified by affinity chromatography. Affinity chromatography exploits the characteristics of protein to bind tightly, non-covalently, and specifically with the ligand molecules for purifying the protein. In this method, ligands specifically bind to a nonreactive analog of the target enzyme's substrate that is linked to an inert matrix by a covalent bond. The target protein in the mixture alone could bind with the specific immobilized ligand.

Moreover, by altering the elution conditions, the target proteins are released from the matrix in a highly purified form. The significant advantage of affinity chromatography is it relies on the biochemical properties of target proteins when compared to other chromatographic techniques which depend on the physicochemical properties of proteins. Therefore, the separation efficiency of affinity chromatography is higher than other chromatographic methods. In this method, the columns are constructed by attaching specific ligands to the matrix. In immune-affinity chromatography, target protein-specific antibody can be bound with the chromatographic matrix to purify the protein. In affinity chromatography, the ligand should possess a high affinity to bind to target protein without denaturing it. The bound protein is eluted using a solution of different ionic strength or pH or using a higher

concentration of free ligand. In metal chelate affinity chromatography, divalent metal ions (Ni^{2+} or Zn^{2+}) are bound to a chromatographic matrix, which is used to purify proteins with chelating metal groups (multiple histidine side chains). His-tag made up of six consecutive His residues can be allowed to bind with the C-terminal or N-terminal of the recombinant protein/polypeptide that can be purified by metal chelate chromatography. After eluting the purified protein by changing the pH, His-tag can be removed using protease, which specifically recognizes (His)x6 sequence from the remaining protein.

2.2 Biophysical and Biochemical Characterization of Protein

Structure of protein is classified into four categories, namely (1) primary structure, the linear sequence of amino acids in peptide chain, (2) secondary structure, organized regions within the peptide chain, (3) tertiary structure, three-dimensional conformation of a polypeptide, and (4) quaternary structure, spatial combinations of multiple peptide chains. Each protein structure has some unique, and it is different from other proteins. Protein structure and its stability play a very important role in determining their function which is confirmed into experimental and computational methods. Protein stability is a net balance of physical energy, which determines whether the protein will be in native, folded, or unfolded conformation. Structure and stability of protein vary from one protein to another. For example, thermophilic proteins maintain their stability at extreme temperatures (70–100 °C), while some proteins are very sensitive and stable only at normal temperatures. Biophysical and biochemical characterization methods are effective in determining the protein structure, thermal stability, particle size, aggregation, degradation, etc.

2.2.1 Biophysical Characterization Methods

Biophysical methods are used to explore the different molecular interactions between protein molecules and solutes. It plays an important role in the development of protein analysis for drug development. Biophysical methods can provide several information including protein interaction and aggregation. Biophysical methods can reveal insights into reaction of protein in solution and its dynamic interaction with each other. These methods are classified into three main categories which includes the following:

- Hydrodynamic methods (analytical ultracentrifugation, viscometry, etc.),
- Thermodynamic methods (light scattering, microcalorimetry, and surface plasma resonance), and
- Spectroscopic methods (fluorescence, circular dichroism (CD) electron paramagnetism).

However, some important methods can be used for biophysical characterization of proteins such as dynamic light scattering (DLS), differential scanning calorimetry (DSC), circular dichroism (CD) spectroscopy, Fourier transform infrared (FTIR) spectroscopy, and isothermal titration calorimetry (ITC).

2.2.2 Dynamic Light Scattering (DLS)

Dynamic light scattering (DLS), also known as photon correlation spectroscopy or quasi-elastic light scattering, is one of the most popular light scattering techniques. DLS is a standardized method for particle size distribution and size analysis of proteins, polymers, nanoparticles, biological cells, gel, and micelles. This technique is very popular to determine the size of particles that range from 0.001 to several microns. In earlier days, DLS was used to determine the hydrodynamic radius of proteins in solution, but recently, size of proteins in solution can be detected using backscattering (Hushcha et al. 2000).

The theoretical background of DLS is based on the Brownian motion of dispersed particles. Those particles are dispersed on liquid, and it will move randomly in all directions. Particles constantly colliding with solvent molecules are defined as Brownian motion. These particle motions are induced by a certain amount of energy, which is transferred during this collision. The smaller particle will be having greater effect due to the Brownian motion because energy transfer is more or less constant on smaller particle. As a result, smaller particles move at a higher speed than larger particles. Hence, the hydrodynamic diameter can be determined by measuring the speed of particles, if we know all other parameters which have an influence on particle movement (Jaramillo-Flores et al. 1998).

2.2.2.1 Applications of DLS

Based on the hydrodynamic sizes, DLS is used for the analysis of

- Aggregation states of proteins,
- Crystal formation,
- Association reactions.

Further studies using DLS precrystallization, assay of aggregation state of insulin (Kadima et al. 1991), lysozyme (Georgalis et al. 1995), and amino-acyl tRNA synthetase (Mikol et al. 1991) prior to crystallization have also been reported. In human serum albumin, a conformational change in small molecules has been reported. These conformational changes are monitored through diffusion coefficients measured by DLS (Hu and Kerppola 2003).

2.2.3 Circular Dichroism (CD) Spectroscopy

CD spectroscopy is a very powerful technique to study secondary structures of macromolecules. This method is widely used to study the biological molecules, their structure and interactions with metal and other molecules. It is also used to determine the secondary structure, protein folding, and interaction between soluble proteins. In genomics projects some proteins are ignored, and these ignored proteins can be certainly solved by CD spectroscopy. A major aspect of CD spectroscopy is that it is the fastest method to detect for protein folding within microsecond time frame. It is also used to detect the protein stability by changes in the composition of secondary structure. CD spectroscopy method is relatively expensive and also could not provide site-specific information (Miles and Wallace 2016; Tripathi 2013).

For proteins, this method will be mainly concerned with the absorption in far ultraviolet (UV) wavelength region between ~240 and 190 nm due to the amide chromophores of the peptide bonds. In wavelength region, there are two kinds of electron transitions responsible for the CD signals, such as

(i) $n \rightarrow \pi*$ transition at around 222 nm,
(ii) $\pi \rightarrow \pi*$ transitions at ~208 and 190 nm (both parallel and perpendicular orientations).

2.2.3.1 Applications of CD

- CD plays a vital role in supplementing higher resolution structural approach of X-ray crystallography & NMR.
- It is a rapid and convenient technique.
- CD is mainly used in determining the structural aspects of protein and as a supplement with other structure defining techniques.
- CD spectroscopy is widely used in protein engineering studies including folding and unfolding characteristics.
- CD is mainly used for studying secondary structure, dynamics folding pathways and interaction of soluble protein that facilitates to analyze stability and conformational changes of protein structure.
- Interaction between thermodynamics of folding and unfolding of proteins and conformation nucleic acids can be evaluated by CD spectroscopy.

2.2.3.2 Advantages of CD

- Uses of very less quantity of sample (200 μL of 0.5 mg/mL solution).
- Non-destructive.
- Relative changes due to the influence of environment on sample like pH, temperature, and denaturant can be monitored accurately.

2.2.4 Differential Scanning Calorimetry (DSC)

Differential scanning calorimetry (DSC) is a very sensitive thermo-analytic technique, which was first developed by M.J. O'Neil and E. Watson of Perkin Elmer in

1960s. DLS is an experimental method to characterize thermal-induced conformational changes in protein and other biological molecules. For example, DLS facilitates to study the different conformations of the proteins, DNA binding, and lipid interactions. DNA drugs and pharmaceutical field with excipient studies have been utilized in DSC. The main role of non-perturbing DSC method is to characterize the stability of a protein, folding mechanism, and thermal behavior of particular compound (Johnson 2013).

2.2.4.1 Applications of DSC

- To determine the melting point of your protein
- Protein folding and thermal stability
- DSC is also used to probe folding free energy surfaces and barrier heights
- Drug discovery and development
- Ligand interaction studies
- Rapid optimization of purification and manufacturing conditions
- Easy, rapid determination of optimum conditions for liquid formulations
- Quick stability indicating assay for target proteins to be used for screening.

2.2.5 Fourier-Transform Infrared (FTIR) Spectroscopy

Fourier transform infrared (FTIR) spectroscopy is one of the major techniques to determine the structural characterization of proteins. Especially protein's secondary structure can be determined by this method which enhance the understanding of protein conformational changes (effects of ligand binding, temperature, pH, and pressure) and stability aggregation of proteins. FTIR is crucial for understanding the structure–function relationship and enzyme kinetics of various proteins to identify the secondary structure of protein. FTIR also enables the analysis of chemical bonds, protein complex, or secondary structure of protein. FTIR measures the amount of light absorbed by sample at each wavelength. The principle of FTIR depends on the fact that most molecules are absorbing the light in infrared region of the electromagnetic spectrum. This absorption resembles the bonds present in molecule. The frequency ranges are measured as wave numbers typically over the range of 4000–600 cm^{-1}.

2.2.5.1 Advantages of FTIR

- Understanding the protein secondary structure.
- FTIR spectroscopy for structural characterization is the lack of dependence on the physical state of the sample.
- Aqueous or organic solutions, hydrated films, nonhomogeneous dispersions, or solids and proteins have been analyzed by FTIR spectroscopy in all of these physical states.
- Rapid collection of data and high light intensity at the detector and in consequence the high signal-to-noise ratio.

2.2.6 Isothermal Titration Calorimetry (ITC)

ITC determines the thermodynamic parameters of all interactions in solution. This method is often used to study the binding of small and macro molecules (Pierce et al. 1999). The main role of this technique is to quantify the non-covalent binding of proteins by both exothermic and endothermic. When two or more proteins interact with one another, the conformational changes in protein or rearrangement result in the absorption or generation of heat. The heat generated in the reaction will be quantified by ITC and also provide a thermodynamic description of the binding interaction and the stoichiometry of binding.

ITC determines the heat evolved during protein–protein interaction and provides stoichiometry (n), binding constant (K_a) as well as the enthalpy of binding (ΔH). This method also enables the change in entropy (ΔS) and the change in Gibbs energy of the system, ΔG, to be acquired from the following:

$$-RT \ln K = \Delta G = \Delta H - T\Delta S$$

where T represents the temperature, R is the universal gas constant.

For efficient protein interaction, the crystal structure of protein complexes necessitates hydrophobicity, charge, and shape complementarity so that the resulting interface is packed like a protein's interior. In theory, the interaction of two proteins can predict the interaction interfaces and the binding mechanism. However, in practice, it is complicated by the dynamic event of the protein and the solvent.

2.2.6.1 Applications of ITC

- To measure the quantification of thermodynamic properties driving protein–protein interactions.
- ITC is widely used for drug discovery.
- Determination of binding specificity and stoichiometry.
- To characterize the thermodynamic properties of protein–protein interactions (Doyle 1997).
- To characterize the binding affinity of ligands for proteins.
- To study the bio-molecular interactions and validation of IC_{50} and EC_{50} values during hit-to-lead.
- To characterize the mechanism of action and measurement of enzyme kinetics.

2.2.7 Biochemical Characterization

To develop protein therapeutics, chemical stability is the major concern due to its impact on both efficacy and safety. Protein "hotspots" are amino acid residues that are subject to various chemical modifications, including deamidation, isomerization, glycosylation, and oxidation. By using these methods, we can modify the stability of protein structure and conformational changes of the protein. Therefore, it is to

determine the protein structure and analyze the stability of the structure. Widely used biochemical analysis techniques are as follows:

- Oxidation
- Deamidation
- Phosphorylation
- Acetylation
- Glycosylation.

2.2.8 Oxidation

Oxidation is a type of chemical reaction that involves the transfer of electrons between two species. Oxidation is the gain of oxygen during the reaction by a molecule, atom, or ion. If oxygen is removed during the reaction, it is called reduction. Both oxidation–reduction reactions are called as a redox reaction which is a vital role in the basic function of life such as combustion, respiration, photosynthesis, and corrosion or rusting. Protein oxidation is defined as the covalent modification of a protein by the direct reactions with reactive oxygen species (ROS) or by indirect reactions with secondary products of oxidative stress (Zhang et al. 2014). In this ROS can cause modification in the side chain of proteins, which leads to structural changes at primary, secondary, and tertiary structure level of proteins. These structural changes may induce conformational and functional modifications of proteins including enzyme activity. Therefore, this method can be used for the characterization of proteins by the biochemical method. Both carbonyl residues and 3-nitrotyrosine residues can be detected by protein carbonyl assays and protein nitration assays, respectively, and are stable markers of oxidative stress.

2.2.9 Deamidation

Deamidation is one of the chemical reactions which are used to remove the functional groups of amide from organic compounds. Moreover, loss of the ammonium group of asparagine and glutamine to form aspartic and glutamic acid, respectively, is called deamidation and is the most commonly occurring posttranslational modification in proteins. Typically, asparagine is converted to aspartic acid or isoaspartic acid. Glutamine is converted to glutamic acid or pyroglutamic acid (5-oxoproline). In proteins, this reaction plays a very crucial role because they may alter the protein structure, function, and stability, and could lead to degradation of protein. In 1990, first Gln deamidation study was published, which exhibited the formation of glutarimide intermediate and deamidation products, α- and γ-glutamate (Glu, E) (Capasso et al. 1991). Protein-based drugs play vital role in drug discovery because it possesses high target specificity with a low side effect (Jia and Sun 2017). They could influence the rate of deamidation in three-dimensional structure (primary sequence, tertiary structure), pH, buffer salts, and solvent properties (Li et al.

2005). Protein deamidation can be analyzed by Peptide mapping using reverse-phase liquid chromatography (RPLC) (Zhen et al. (2018). Mass spectrometry can be used to characterize the deamidation states of a protein. This technique is helpful for the analysis of deamidation due to its high speed, specificity, and sensitivity (Hao et al. 2017). The main factor of deamidation reactions is that it has been found to limit the useful lifetime of proteins (Clarke 2003). These reactions continue very rapidly at high pH (>10) and temperature.

2.2.10 Phosphorylation

Protein phosphorylation is a reversible posttranslational modification catalyzed by protein kinases. It was first reported in 1906 by Phoebus Levene with the discovery of phosphorylated vitellin. Phosphorylation is one of the chemical reactions, the addition of phosphate group to a molecule. It plays a crucial role in the mechanism of protein and function of the enzyme. Protein phosphorylation can be regulated to cellular functions, and it may change the structure of the protein. The main mechanism of phosphorylation is the transfer of a phosphate group to the amino acids side chain (Erlandsen and Stevens 1999). After the reactions, it will activate or deactivate, or modify its function. Approximately 13,000 human proteins have sites that are phosphorylated. Serine, threonine, and tyrosine are the most commonly phosphorylated amino acids. Thirty percent of human proteins can be phosphorylated, and abnormal phosphorylation was now identified as a cause of human disease. Abnormal phosphorylation causes a number of diseases such as Alzheimer's disease, Parkinson's disease, and other degenerative disorders. Protein phosphorylations lead to changes in protein stability, function, and localization (Johnson and Lewis 2001). Some important methods can be used for the detection of protein phosphorylation such as western blotting, ELISA, quantitative mass spectroscopy, multi-analyte profiling, and intracellular flow cytometry. The most domain of phosphorylated and unphosphorylated crystal structure was reported (Kobe et al. 1999). Some of the main functions of protein phosphorylation are as follows:

- Degradations of proteins
- Gene transcription
- Cellular proliferation
- Signal transductions
- Regulating pathways
- Regulations of enzymes
- Protein–protein interactions
- protein localization
- To activate the enzyme activity.

2.2.11 Acetylation

Acetylation is the introduction of an acetyl functional group to the lysine amino acids of histone tail. It will provide binding sites for a number of proteins. Acetylation is a typical metabolic reaction with thiol groups, hydroxyl groups, and often amino groups (Christensen et al. 2019). Acetylation can alter the stability of proteins, expression of genes, and protein–protein interactions. The removal of an acetyl group is called deacetylation. Acetylation was first detected in histones. The acetyl group is transferred from acetyl-coenzyme A catalyzed by acetyltransferases (Cunin et al. 1986). In many cellular processes, acetylation and deacetylation are very crucial methods. Two kinds of acetylations of amino groups of proteins can occur, lysine acetylation and N-terminal protein acetylation. The main function of acetylations is the stability of proteins, protein localization, synthesis of protein, metabolism, and apoptosis. Acetylation can be detected by a mass spectrometer (Brown et al. 2017; Kentache et al. 2016; Ouidir et al. 2015), so that, we can easily predict the protein stability, also to detect the function of the protein used by acetylation.

2.2.12 Glycosylation

The addition of carbohydrate moiety to a protein molecule is called protein glycosylation to occur in protein biosynthesis (Fig. 2.1). Glycosylation is one of the enzymatic processes. This method is one of the most common posttranslational modifications (PTM) because posttranslational modifications are very essential for the stability of proteins, protein–protein interactions, cascade amplifications, localizations of protein, biochemical activity. Here, the role of PTM is a modification of amino acids.

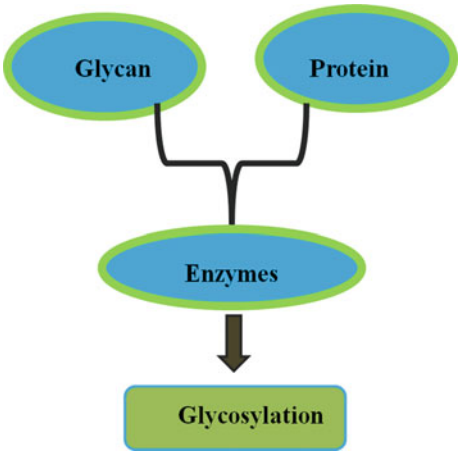


Fig. 2.1 Overview of glycosylation

Protein glycosylation takes place in the endoplasmic reticulum (ER) and golgi body. It mainly helps in the appropriate folding of the protein and maintaining the protein structure. For example, some proteins do not fold properly unless they are glycosylated and cannot be exported from the ER. Some glycoproteins act as a carrier, it will directly bind to the molecules, for example for ceruloplasmin.

The main purpose of glycosylations is cell–cell adhesion, to avoid degrades, folding, and stability. Over 40 disorders of glycosylations have been reported in humans. With the help of amino acid modifications, we could analyze protein structure stability, easy to predict the function of the protein.

2.3 Structure Determination

Proteins are marvel molecules, omnipresent in all living organisms that essentially perform all functions required for survival. Whatever functions performed in any living organism, such as transport, storage, mechanical support, movement, and growth, are accomplished by this phenomenal molecule termed as proteins. To achieve this, we have to reveal the structure of these biomolecules which will give insights about its function. Proteins are biopolymers that consist of 20 different amino acids determined by the gene that codes it (monomer) linked by peptide bond which eventually folds into three-dimensional structures by inter-atomic interactions. However, the structure of proteins can be described into four different levels (Desai et al. 2010; Bencharit and Border 2012).

The primary structure is the linear sequence of specific amino acids arrangement of the protein. The secondary structure extends to a short range of sequences formed because of the local interacting hydrogen bonds between backbone C–Cα, mostly forms α-helices or β-sheets. Tertiary structure is the three-dimensional arrangement formed between side chains of amino acids resulting in different conformations with secondary structures like α-helices and β-sheets connected by disordered regions termed as turns or loops. Quaternary structures are complete spatial arrangements of two or more subunits within a protein molecule (Walsh et al. 2015). The different levels of organization of protein structure can be determined using various methods such as primary structure, i.e., sequence of any given protein can be found by Edman degradation method and mass spectrometry. The secondary structure can be determined by circular dichroism and FTIR. The three-dimensional structure of a protein (tertiary/quaternary) can be determined by either experimental methods like X-ray crystallography, NMR, Cryo-EM, and neutron diffraction or can be predicted by computational methods like homology modeling, ab-initio, and threading methods (Walsh et al. 2015). Each method has its own merits over the other with certain limitations, whereas the most popular and widely used method for experimental determination of protein structure is X-ray crystallography, and with over 89% of structures deposited in Protein Data Bank were solved by this method, followed by NMR over 8%, Cryo-EM about 2.5%, and the remaining 0.5% structures were solved using other methods of 156,365 structures currently deposited in PDB

(Burley et al. 2019). In this chapter, the basics, principle, and structural determination of protein by various methods will be described briefly.

2.3.1 X-Ray Crystallography

In simple terms, X-ray crystallography means the investigation of crystals by X-ray. The prerequisite for the determination of protein structure using X-ray crystallography is to obtain a crystal. The crystallization of proteins is difficult as the actual physics behind it is not well understood; lots of factors affect crystallization processes like pH, temperature, vibration, protein and precipitant concentration, protein purity, and flexibility (Parker 2003). Protein crystals are obtained by supersaturation of protein solution either by increased concentration of precipitant or by protein. X-rays are produced when accelerated electron (synchrotron) or electron hits an anode metal, the crystal is mounted in a goniometer, a device that rotates 360° so that X-rays can be exposed throughout the crystals. When X-ray is exposed to crystals, they are scattered by the electron present in an atom forming diffraction patterns that contain structural information of the molecule (Smyth and Martin 2000). However, detection of diffraction by single molecule will be too weak, but the crystals which are orderly fashioned repeats make it possible to detect. The basis for this experiment is laid by Laue and W. L. Bragg, named as Bragg's law, accordingly X-ray diffraction is considered similar to the reflection of atoms in a plane of crystals. Constructive coherent interference should occur for a diffraction pattern to be obtained which appears only when the given equation is satisfied,

$$n\lambda = 2d \sin\theta$$

where λ is the wavelength of X-ray, d is the inter plane distance, and θ is the angle between the incident ray and planes.

The scattered X-rays are recorded by CCD (charged coupled device) detectors and the diffraction intensity data were collected, which may be influenced by several factors such as quality of the crystal, crystal symmetry, X-ray intensity, and detectors efficiency. This collected data is processed using modern software for crystal system identification, indexing, measurement of intensities and scaling factor to get electron density maps. The main bottleneck of crystallography is phase information, that is not obtained, thus this has to be fetched using various methods like molecular replacement where experimentally solved homologous protein structure's phase information is used, isomorphous replacement data is collected in native form and with heavy atom soaked form, when this signal is weak then multiple heavy atoms can be soaked which would eventually result in multiple heavy atom signal and thereby phase information can be fetched. With the obtained amplitude and phase information, electron density maps are constructed using a fast Fourier transform method. This will give the three-dimensional contours from which the structure will be build and further the quality of the structure can be found by the R-factor. Once

the model of the structure is predicted, further refinement by computer programs like Refmac will be performed to get better quality structures (Su et al. 2015).

Though X-ray crystallography fetches the highest atomic resolution of protein structures, there are certain limitations in it such as the protein needs to be soluble, crystallizable, and is required in large quantity. But bigger proteins are difficult to express and crystallize, and therefore the target proteins that are comparatively larger than 250 kDa are hard to be solved by this method. Though X-ray crystallography fetches the highest atomic resolution of protein structures, there are certain limitations in it such as the protein needs to be soluble, crystallizable, and is required in large quantity. But bigger proteins are difficult to express and crystallize, and therefore the target proteins that are comparatively larger than 250 kDa are hard to be solved by this method.

2.3.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a technique used to measure the absorbance of radio frequency (RF) radiation by certain nuclei like H^1, C^{13}, N^{15} when exposed under high magnetic fields. This works by the principle that certain molecules are magnetic and their nucleus spin in the presence of magnetic field. When external magnetic fields are applied, protons exhibit certain spinning states (α or β forms), and the difference between these states is related to the strength of the applied magnetic field. The spin state can be shifted from α form to excited β form by applying the appropriate pulse frequency of electromagnetic radiation specific to atoms which is termed as resonance. This resonance spectrum can be obtained by varying electromagnetic radiation or magnetic field by keeping either one as constant. To determine the structure of a protein using NMR, it should first be labeled with C^{13} and N^{15} atoms that will enable it to study the structure which can be achieved by recombinant expression of protein in C^{13}-labeled glucose and N^{15}-labeled ammonium chloride as sole carbon source and nitrogen source, respectively, which results in radiolabeled protein (Sugiki et al. 2017).

The structure determination by NMR is performed by two data sets, first is the measurement of nuclear magnetic resonance by isotopic labeled molecule such as C^{13}, N^{15}, and proton that has specific resonance when exposed in electromagnetic radiation. Resonance differ based on the nuclei, frequency, and the electromagnetic environments in three-dimensional structure of the protein which will be assigned to a specific amino acid that will be found using various experiments. The second data set has information about internuclear distances and the position of the atom that can be found by the NOE (Nuclear Overhauser effect), i.e., magnetization effect is shifted from one nuclei to other which is less than 5 Å apart. This provides the position of an atom relative to one another in three-dimensional space which can be used to reconstruct the structure of the protein. Several factors such as inadequate data set, approximate distance information and necessity of proteins in solution state affect the determination of protein structure by NMR (Wider 2005).

The proteins that cannot be crystallized but are soluble and are relatively small in size (2–40 kDa) can be solved using NMR, and ensembles of protein structure are determined in solution and therefore the dynamical behavior of the protein can be studied. The resolution of protein structure solved in NMR is relatively low compared to X-ray crystallography where essential information like active site interactions are difficult to obtain.

2.3.3 Cryo-electron Microscopy

Cryo-electron microscope is similar to the light microscope in principle but the difference is instead of light as radiation source with a higher wavelength of 4000–7000 Å, electron that has a wavelength of 0.02 Å (300 kV) is used which enables interaction with molecules in a greater extent compared to light. It is a widely used technique in visualizing cell architecture and complex structures such as protein assemblies at a molecular resolution which cannot be achieved in any other method, and protein as large as several mega Daltons structure can be solved. With the latest advancements in computational and imaging techniques, higher resolution (1.8 Å resolution) of the AAV2 capsid variant (3.9 MDa) has been recently solved (Tan et al. 2018).

First, the macromolecular sample in solution state is applied onto a carbon-based film with EM grid spread across to have even molecule distribution by special treatment of grid through ion bombardment, creating a hydrophilic surface to hold protein and are frozen using liquid nitrogen or liquid helium to form amorphous ice for the protection of samples from radiation damage. This vitrified water (ice) is used as a substratum to hold the molecules intact and sandwiched. The frozen grids are transferred to cryo-transfer holders where imaging is done with a low electron dose to reduce radiation damage. Then, several thousand images were recorded by tilting the specimen that reduces the need for rotation, with the molecules in different orientations using an advanced detector. The images are collected in defocused condition to typically increase the phase contrast that will reduce the noise-to-signal ratio.

The obtained cryo-EM micrographs are computationally reconstructed to get a three-dimensional model upon performing the Fourier transform of projection. Once the data reconstruction is done beyond 4 Å resolution, certain tweaking like the de novo model by various refinement programs will result in higher resolution data (Lyumkis 2019).

2.3.4 Prediction of Structure Using Computational Methods

Proteins structures that could not be found using the various experimental techniques discussed earlier either due to time constraints or because of experimental challenges can be predicted by different computational approaches. These structures can be compared at par with experimentally solved structures but with a certain degree of

variability. Ab initio prediction methods can be used to predict the structure directly from sequence without any reference to the structural template using quantum mechanics. Small proteins with a sequence length of about 100–150 residues and with no availability of structural template, this method would be used. But the accuracy is one of the major problems in this method; moreover, there are certain assumptions like neglecting the side chains and considering peptide bonds as a plane (Venkatesan et al. 2013).

2.3.5 Homology Modelling

Homology modeling is a method of protein structure prediction which is based on the principle that protein sequences are evolutionarily related, similar protein sequences have more conserved structures. Many protein structures deposited in PDB share similar folds, in turn, lots of protein structures can be solved using this method. The only criterion is the protein should share at least 40% sequence identity with the template when it is aligned pairwise. First, the homologous protein template is identified from PDB based on the sequence match. Then the pairwise alignment is carried out between the query and template protein, using which an initial model is built with insertion, deletion, residue substitution from template followed by optimization of model. This is a widely used methodology to predict the protein structure, provided the experimentally solved homologous template structure should be available (Nikolaev et al. 2018).

2.3.6 Threading Methods

Threading is similar to that of homology modeling, but the difference is it does not require having a homologous sequence. Here the protein sequence is aligned to the backbone structure with an evaluation parameter such as a quasi-energy function to assess the sequence–structure compatibility. The concept behind it is that the number of unique folds is lesser than the number of proteins in nature. Protein domains that are a smaller subunit of a protein have folds independent of the protein. The parameters in this method not only consider the amino acids but also their position with the other properties like secondary structure, solvent accessibility, and adjacent amino acids spatially (Shehu et al. 2016).

2.4 Role of Molecular Interactions in Structural Stability

The complete valence shell of atoms and molecules are tightly packed to one another to form the defined structure. Each atom of a molecule is associated with a specific charge (positive or negative) to attract or repel the atoms defined as electrostatic interactions. These molecular interactions are an important parameter in the diverse field of life science. For example, it helps to understand protein folding, drug

designing, and material science, nanoscience, etc. Intermolecular interaction or non-covalent interaction that forms interactions between two or more molecules or atoms is called intra-molecular interactions (Reece et al. 2011). Non-covalent interactions are differing from covalent interaction, which does not share electrons rather it forms force between the atoms. Usually, bonds between the atoms hold the molecule in particular conformation in their environment which is important to restrain the stable structure. The atoms like nitrogen, oxygen, iodine, and hydrogen are present in the two-atom molecule at room temperature. Under high pressure and temperature, Nitrogen atoms lose their interaction thereby deform the native structure called a chemical reaction.

The amount of energy consumed to break the bonds between the atoms is defined as bond enthalpies. For instance, breaking the bonds of water molecule requires 493 kJ/mol energy; likewise, all the atoms do break the bond to some extent. Bonds in the biological molecular are the same in DNA, RNA, and protein unfolding and are not involved in breaking the bonds to deform the structure. The enthalpy energies of the molecule can be varied according to the environmental condition. Mostly, the molecular interactions in structure give stability to attain the biological function of the complexes. The protein forms particular fold-like structure and stabilized when it forms globular fold. DNA molecules always have a double standard helix for biological function. It maintains the structure due to the molecular interaction between the helices. As such, protein binds with a DNA groove for complex formation which is the cause of molecular interaction (electrostatic interaction) between the two complexes. In the biological assembly of protein–DNA complex, proteins recognize and bind to the DNA groove. In contrast, when the complexes unfold or disassemble the structure, the surface of the protein is surrounded by the water molecule; therefore, the molecular interactions occur with the surrounded water surface. Hence the nature of the protein or DNA may differ from the native fold of the complexes.

The term molecular interaction was first coined by Johannes Diderik van der Waals. The van der Waals interaction defines the attractive or repulsive force of the atom (Roth et al. 1996). Molecular interactions are electrostatic in nature which is charged surface of the molecule. The structural stability of protein, DNA, or RNA is relying upon the molecular interaction. In the drug discovery process, physicochemical properties and geometry of the ligand molecule are important for binding to the specific binding site for inhibitory activity. The covalent single bond of the compound comprises of 80–100 Kcal/mol energy and few non-covalent interactions (Andrews et al. 1984). Usually, energies of salt Bridge, H-Bond, hydrophobic and aromatic interactions consist of ~2, ~1, ~0.7, and ~ 1–3 kcal/mol, respectively. Also, we can measure 25-fold increased affinity for single ionic interactions when $dG = -2.303\ RT \log K$; likewise, a hydrogen bond produces six-fold increased energy and 3.5-fold increased energy for the methyl group (−CH3). Even a steric clash between atoms induces a remarkable impact on the binding affinity, which was simply explained by the Lennard-Jones potential. The term molecular interaction comprises of several types which are (1) electrostatic interactions, (2) salt bridges, (3) hydrogen bonds, (4) short-range repulsion, (5) ion pairs in proteins, (6) dipolar

Interactions, (7) dipole–dipole interactions, (8) dipole-induced dipole interactions, (9) charge–dipole interactions, (10) fluctuating dipolar interactions, (11) cation-Π interactions, and (12) the Lennard-Jones potential and van der Waals radii.

2.4.1 Electrostatic Interactions for Biological Functions

In protein, the ionizable side chains of amino acids such as Asp, Glu, His, Lys, and Arg residues are important to define the protein property. Charges of these amino acids can modulate by changing the pH, thereby it affects the protein folding nature. Also, phosphorylation and de-phosphorylation alter the charges of Ser, Thr, and Tyr residues which involves in the protein–protein interactions for signaling mechanism. Posttranslation includes charge modification, acetylation of lysine, enhancing the strength of the protein–DNA association for the complex formation. In transcriptional regulation, the protein and DNA molecules have their charged surface binding strongly due to the electrostatic interactions. Charges of the proteins are vital for the binding of ligand and metal ions in the specific site of protein through the transmembrane channel. In addition to that, the polar residues consisting of partial charges on the structure are involved in partial electrostatic interaction between the charged residues for protein folding, binding recognition, and condensation. Charged residues are significantly differing from the non-polar residues as the hydrophobic interaction of the residues provides the driven force to enhance the structural stability of the protein. The charged residues consist of either a positive or negative charged surface that binds with the oppositely charged residues. Charged residues interact with polar residues and make specific hydrogen bond interactions. Most significantly, charge–charge interaction forms stable interaction even at a distance of 5–10 Å which facilitates binding of the small molecules in the protein for complex formation (Zhou and Pang 2018).

2.4.2 Salt Bridges in Protein Structure

Salt bridge is a type of non-covalent interactions, typically forms interaction between two ionized sites of the amino acids in the protein. In general, the side chain of opposite charge of amino acids (Glu− or Asp− vs. Arg+ or Lys+) forms salt bridges. Also, amino acid side chains in protein can form salt bridges with ionized ligand, such as K^+ and SO_4^{2+}. It also contributes to the structural stability of the native protein. Mostly, the salt bridges are believed to exist in the center of charged atoms with a distance of 4 Å. The proteins of thermophiles have additional salt bridge than the protein of *mesophiles* organism. It gives the structural stability to protein and resistance to live in the thermal environments. Salt bridges consist of two main components, namely hydrogen bonds and ionic interactions. Hydrogen bonds are very crucial for the ligand binding, protein–protein interaction, and enzymatic activity. Ion pairing interactions are one of the most crucial interactions and involved in the structural stability of the proteins. Also, the proton donor and acceptor of

atoms form hydrogen bond interaction with an average distance of 3 Å which shows that the distance in hydrogen bond could differ from the ion-pair interactions. Mostly, non-covalent interactions are weak interaction, but it contributes less to structural stability in protein conformation (Gromiha 2010).

2.4.3 Hydrogen Bonds

Hydrogen bonds form the interactions with and within the two atoms of molecules. It contributes to the configuration of protein structure folding and stability. Yet the researchers are unsure about the nature and importance of hydrogen bonds in the protein structure completely. In hydrogen bond formation, the hydrogen atom form interaction with electronegative atoms such as O_2, Cl_2, and F. The interaction is very strong compared with van der Waals interaction but weaker than covalent bond interaction. In general, hydrogen bonds found in the DNA base pair of two strands give structural geometry and stability. In protein, each hydrogen bond in the structure gives 5 kcal mol^{-1} of stability in a unique configuration.

2.4.4 Short-Range Repulsion

The forces of attraction and repulsion of two molecules are defined as intermolecular interaction in which short-range force occurs between the center of two molecules with a distance of 3 Å. Mostly, short-range forces likely to be repulsive since the two or more atoms present near to the orbitals on the surface of the atom are of very short range. It is very important for protein structure stability because it prevents the residual collapse in maintaining the structural folds (Feke et al. 1984).

2.4.5 Ion Pairs in Proteins

Ion pairs generally consist of positive and negatively charged particles. Adequate energy given to neutral atoms produce charged particles of ion pairs that results in occurrence of dissociation in the oppositely charged molecule. Ion pairs are vital interaction between the oppositely charged amino acids for the stabilization of protein structure. They are simply classified as a class of salt bridges that forms the single conformer hydrogen bonded ion pair in the crystal structure. In solution, ion pair is important for stabilizing α-helix in the domain which consists of charged residues (Arg, Glu, and Lys) in the turn of the helix to form interaction (Batchelor et al. 2019).

2.4.6 Dipole–Dipole Interactions

Dipole forms interactions with the unequal pair of electrons between atoms or molecules. Dipole–dipole interactions form when two dipolar molecules interact with one another. The partially charged negative region of polar molecule attracts the partially charged positive portion of another polar molecule. In protein structure, dipole–dipole interactions are playing a major role in stabilizing the helical structure. This interaction generally dictates to adapt a distinct conformational change in the protein structure for binding with other proteins or peptides. Also, dipole–dipole interactions are important for protein–protein complex formation.

2.4.7 Pi–Pi (Π–Π Stacking) Interactions

Pi–pi stacking interactions are the type of non-covalent interaction formed between the aromatic ring of molecules or amino acids. The residues tryptophan, tyrosine, and phenylalanine consist of an aromatic ring which may form pi–pi stacking interaction with other aromatic containing molecules or compounds. Pi–pi staking interactions are important in the structural stability and molecular recognition of protein–protein and protein–ligand for biological mechanism. The aromatic ring consisting of ligand molecule will approach the active site residues of protein via pi–pi stacking interaction which later binds with the pocket with high affinity.

2.5 Importance of Physical Properties and Its Influence on Structural Stability

Proteins are complex macromolecules found in all living systems and serve a crucial role in the biological process. For example, they catalyze thousands of chemical reactions as a catalyst, they are messengers to transport ions, electrons, and small molecules, and they also regulate gene expression besides protecting cells from foreign particles in the form of the antibody. Indeed, proteins play a dynamic role in all living systems. They are polymers of amino acids constructed from 20 different amino acids and linked by amide groups known as a peptide bond. Each amino acid consists of alpha carbon bonded to four structural groups such as hydrogen atom, carboxyl group, amino group, and a variable R group (Fig. 2.2). The R group comprises different chemical properties and it varied from one amino acid to another.

In general, aminoacids are regulating protein metabolism. Similarly, the structure and function of the protein are determined by a linear sequence of amino acids. Protein has four hierarchical structural organizations: primary, secondary, tertiary, and quaternary structures. In primary structure, amino acids are linked with a peptide bond, while in alpha-helix and beta-sheet forms of secondary structure, the repeating pattern of a hydrogen bond is formed between the backbone atoms. The tertiary structure of the protein is a three-dimensional structure enabled by non-covalent interactions such as hydrophobic bonds, electrostatic bonds, and van der Waals

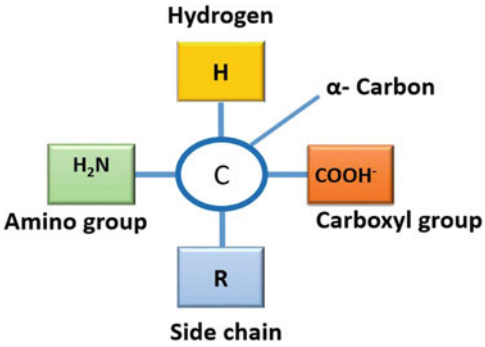


Fig. 2.2 Amino acid structure

forces. Quaternary structure designates interaction between one or more polypeptides to regulate the protein function using the same forces. Every protein possesses its unique properties to carry out its function. To understand the diversity of protein function, it is necessary to look for the physicochemical properties of the protein.

2.5.1 Physical Properties of Proteins

Every protein has specific physical properties to maintain their structure and function. Moreover, molecular forces and interactions in the protein structure determine the stability of the protein. Similarly, covalent bond, electrostatic interaction, hydrogen bond, disulfide bond, and van der Waals interaction in the protein structure regulate their proper folding and functional mechanisms. Changes occur in the properties due to environmental stress like temperature, pH, and osmotic pressure which affects the molecular interaction, structural stability, and function of the protein. In consequence, physical properties are an important tool for maintaining the structural stability and function of the protein. Following factors are important in predicting the physical properties of proteins:

2.5.2 Shape and Size

Generally, proteins are of two distinct shapes: spherical globular protein and strand-like fibrous protein. Globular proteins are soluble; hydrophobic groups are packed in the interior away from the water, and only hydrophilic groups are exposed to the outside which is the reason for its soluble nature. They are folded into a ball-like structure (e.g., hemoglobin). Fibrous proteins are thread-like structures, present in the form of wires. They are insoluble and the peptide chains are bound together by strong disulfide bond making the protein more stable (e.g., collagen and keratin). Globular proteins have multiple functions, and they catalyze the metabolic reaction

and cellular messengers to transport oxygen around the body. Fibrous proteins are involved in the structural function. The globular nature of the protein is determined by ultracentrifugation and dynamic light scattering techniques.

2.5.3 Molecular Weight

The molecular weight of the proteins is varied from smaller 5000 Da to as large as several million daltons. The molecular weight of protein depends on the number of amino acids present (average weight of one amino acid is 110 da). Hence, the amino acid composition is one of the parameters used to calculate the molecular weight of proteins.

2.5.4 Denaturation

Denaturation of protein includes the destruction of secondary and tertiary structures of protein without affecting the primary structure. During the denaturation process, some physical, biological, and chemical alterations occur in protein structure and disrupting their functional activity**.** Moreover, a variety of physical conditions and chemical reagents can cause denaturation by affecting the bonding interaction of secondary, tertiary, and quaternary structures. The main observation of the denaturation process is precipitation and coagulation.

2.5.4.1 Effect of Physical Condition

A physical condition such as heat, UV radiation, and repeated freezing causes denaturation in protein structure by disrupting the hydrogen bond between the amide groups in secondary structure and hydrophobic interaction, salt bridges, disulfide bonds in tertiary structure (Matsuura et al. 2015). Usually, heat increases kinetic energy which modulates the reaction and causes molecules to vibrate so rapidly that the bonds are disrupted.

2.5.4.2 Effect of Chemicals

Chemical reagents include organic solvents (ethyl alcohol) and acid (tannic acid), and heavy metals (mercury, lead, and silver) cause denaturation which affects hydrogen bonds, salt bridges, ionic bands, and sulfide linkages in the protein structure. In some cases, denatured proteins are renatured under suitable conditions and enough time, and it may again exhibit the biological activity.

2.5.5 Amphoteric Nature

Protein also exhibits amphoteric nature similar to that of aminoacids. They react both with alkali and acid in solution. Every protein has a specific pH, and at this pH protein has no charge and cannot move toward either a positive or negative electrode

called isoelectric pH or isoelectric point (pI). The charge of the protein is influenced by pH and the increase or decrease in pH above isoelectric point leads to negative or positive charges in proteins that move towards anode or cathode respectively.

2.5.6 Solubility

Protein solubility is an important physical property and provides excellent information about their function. The solubility of the protein primarily depends on amino acid composition, molecular weight, and polarity (polar and non-polar) of the amino acid surface. Based on these properties, protein solubility is differing from one to another. The solubility of the protein is influenced by various extrinsic factors including pH, ionic strength, and temperature (Trevino et al. 2008).

2.5.7 Effect of pH

pH is an important factor in determining the soluble nature of the protein. Solubility highly depends on the electrostatic repulsion and hydrophobic interactions between the molecules. At the isoelectric pH, the solubility of protein is very low. Protein solubility is enhanced by increasing or decreasing the pH of the solution.

2.5.8 Ionic Strength and Temperature

The solubility of the protein also depends on salt concentration. In lower salt concentration, protein solubility increases (salting in) and after reaching a maximum of solubility, it decreases (salting out). Increasing the salt concentration higher than 1 M may cause decreased solubility. Temperature also influences the solubility of the protein. Every protein has a specific temperature, sometimes increasing the temperature causes insoluble precipitate form.

2.5.9 Computational Tools Used to Analyze the Physical Properties of Protein

Various computational tools used to analyze the physical properties of protein are as follows:

UniProt	http://www.uniprot.org/
PsychoProt	http://psychoprot.epfl.ch/
Variability from alignment	http://psychoprot.epfl.ch/aln2data.html
ProtParam	http://web.expasy.org/protparam/
ProtScale	http://web.expasy.org/protscale/
MultiProtScale	http://lucianoabriata.altervista.org

2.6 Conformational Analysis to Understand the Structural Folding

Proteins are macromolecules, made up of a sequence of amino acids that play a fundamental role in the spatial arrangement of protein structures. Each amino acid in the protein is covalently attached by the peptide bonds to form a polypeptide chain mediated by hydrolysis. The linear form of the amino acid sequence is termed as the primary structure. The amino acid of proteins consists of different properties, and it highly depends on their side chains. Most of the amino acid groups are hydrophobic in nature, and some possess positively and negatively charged side chains. The side chains of charged amino acids are capable of forming ionic interactions and the polar amino acids form hydrogen bonds within the proteins. Also, the hydrophobic nature of side chains interacts with other residues through van der Waals interactions. Significantly, Cysteine residues form covalent interactions to maintain the structural conformation and provide stability to the protein structure. These chemical features of the amino acids interact with other amino acids to form a particular conformation that plays a major role in the biological function of the protein. Precisely, the dihedral angle of the amino acids and the molecular interactions of the proteins propose the significant folds of the protein. The primary structure of protein typically comprises amino acid sequences that drive the protein folding and bond formation for three-dimensional structures of the protein. The secondary structures of proteins are alpha-helix and beta-sheets that form hydrogen bonds with carboxyl and amino groups of the amino acids to form stable folds. These ensembles form particular folds called tertiary structure. Finally, the quaternary structures are formed by multiple subunits of tertiary or secondary folded structures. These properties enable the protein to form particular folds which are energetically minimized structures and are functionally active. To attain a particular structure, protein must undergo several conformational changes to reach the final form. Therefore, non-covalent interactions of amino acids and chemical force of the proteins in the defined environment give defined shape and structural stability to proteins. While protein folding not only relies upon the amino acid strung together, it also depends on the interaction of macromolecules present in the cytoplasm to the partially folded protein, thereby the inappropriate association of protein can attain proper folding to form protein aggregation. This protein called molecular chaperones is essential in the cells to prevent the inappropriate association of protein.

2.6.1 Conformational Analysis of Protein Folding

Each molecule in the living system consists of specific kinetic and thermodynamic properties which depend on the conformation, energies, arrangements, and chemical features of the molecules. Moreover, proteins must attain their unique structural folds via unpredictable ways and conformations (Nachiappan et al. 2019). If the protein folding process is successful, the health of biological systems will be maintained and the absence of this will induce disease. Although, the protein folding

process and the way it attains the folding is quite puzzling and cannot be fully understood by biophysical and biological methods. In the early stages, researchers assumed that protein folds are mediated by the intermediates of distinct pathways (Englander and Mayne 2014). Researchers predicted that the protein folds are formed without any external effects on the protein and it randomly selects their native structure by a number of structural options (Anfinsen et al. 1961). Fundamental differences in protein folding have been evident by protein biophysics methods (Plotkin and Onuchic 2002) by experimentally solving the structures in the unfolded structure to the native structure. But it is quite difficult to analyze the unfolded form of protein to native structure properties. Usually, crystallography and NMR methods cannot define the partial structure rapidly and therefore researchers are forced to select spectroscopic methods such as fluorescence, circular dichroism, and infrared spectroscopy. These methods can explain the kinetic folding of the proteins, but it does not show any structural information. Therefore, researchers found an alternative method to avoid the misconception of protein folding. The theory-based computer simulation is one such powerful method that tracks the protein-folding process by applying biophysical parameters and extensive force field consisting of chemical parameters (Withers 2013). The existing experimental studies have been very difficult to understand the conformational transition of protein. To understand the protein folding and aggregation, it is important to analyze the atomic jiggling and wiggling of the molecules, thereby protein motion, conformation, and allosteric regulatory mechanisms that can be explained using molecular dynamics simulation studies. The intermediates and their pathways involved in protein folding can be explained in three ways: (i) during the kinetic folding, the protein intermediates structures attain the significant occupancy, (ii) the excited state of protein conformation at their equilibrium Boltzmann level high in the free energy landscape above then native protein, and (iii) modification of molten globular protein becomes a higher energy state to lower energy state by destabilizing the native protein. In computation, graph theory analysis provides a significant conformation analysis of protein structure (Hubner et al. 2006). Each residue in protein structure is considered as a node and connecting two nodes with a particular distance (r) is constructed using graph theory. This method explains the significant residual contribution for protein conformation and energy to confer the structural changes that can be analyzed (Englander and Mayne 2014; Anfinsen et al. 1961).

2.7 Conclusions

An essential step for understanding the properties of cells in an organism is to map or predict the structure, stability, functions, and interactions of its proteins completely and accurately by both experimental methods and computational techniques. Proteins can be purified by differences in solubility by salting out process. Proteins can also be purified by ultracentrifugation methods based on differences in size.

Proteins are purified by chromatographic techniques based on size (gel filtration chromatography), charge (ion-exchange chromatography), polarity (hydrophobic interactions), and binding property (affinity chromatography). In all chromatographic methods, the efficiency of protein purification depends on the pH and salt concentration. Electrophoretic methods also purify proteins based on size (SDS-PAGE) and charge (PAGE) that can be used to resolve many proteins. Hydrodynamics (analytical ultracentrifugation, viscometry), thermodynamics (light scattering, microcalorimetry, and surface plasma resonance), and spectroscopic methods (fluorescence, circular dichroism (CD) electron paramagnetism) are the biophysical methods used to explore the molecular interactions between proteins and solutes. Oxidation, deamidation, phosphorylation, acetylation, and glycosylation are widely used biochemical techniques to modify the stability and conformations of the protein. The 3D structure or tertiary/quaternary structure of proteins can be determined by in-vitro methods such as X-ray crystallography, NMR, Cryo-EM, and neutron diffraction or can be predicted by in-silico methods like homology modeling, ab-initio, and threading methods. Structural stability of protein, DNA, or RNA rely on the molecular interactions, namely electrostatic interactions, salt bridges, hydrogen bonds, and short-range repulsion, ion pairs in proteins, dipolar interactions, dipole–dipole interactions, dipole-induced dipole interactions, charge–dipole interactions, fluctuating dipolar interactions, cation-Π interactions, and van der Waals radii. Study of physicochemical properties of proteins, especially shapes, size, molecular weight, denaturation, solubility, amphoteric nature, the effect of pH, ionic strength, and temperature, is essential to understand the diversity of protein function. Several computational tools including UniProt, PsychoProt, Variability from alignment, ProtParam, ProtScale, and MultiProtScale are used to analyze the physical properties of the protein. The intermediates of protein folding and their pathways can be explained during the kinetic folding or the excited state of protein conformation at their equilibrium or during modification of a molten globular protein. Graph theory is used in computational methods to analyze the conformation of protein structure. Unveiling the structure and stability of proteins within the cell by the specified methods can enhance the understanding of disease at the molecular level and augment the development of potent therapeutic or drug molecules.

Acknowledgment J.J. and his group thank the DST INDO-TAIWAN (GITA/DST/TWN/P-86/2019 dated: 04/03/2020), Board of Research in Nuclear Sciences (BRNS) (35/14/ 02/2018 BRNS/35009), Indian Council for Medical Research (ICMR) (No. BIC/12(07)/2015), DST-Science and Engineering Research Board (SERB) (No. EMR/2016/000498), UGC Research Award (No. F. 30-32/2016(SA-II) Dt.18.04.2016), DST-Fund for Improvement of S&T Infrastructure in Universities & Higher Educational Institutions (FIST) (SR/FST/LSI-667/2016) (C), DST-Promotion of University Research and Scientific Excellence (PURSE) (No. SR/PURSE Phase 2/38 (G), 2017 and MHRD-RUSA 2.0, New Delhi (F.24-51/2014-U, Policy (TNMulti-Gen), Dept. of Edn., Govt. of India, Dt.09.10.2018).

References

Andrews PR, Craik DJ, Martin JL (1984) Functional group contributions to drug-receptor interactions. J Med Chem 27(12):1648–1657

Anfinsen CB, Haber E, Sela M, White FH Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A 47(9):1309–1320

Batchelor M, Wolny M, Baker EG, Paci E, Kalverda AP, Peckham M (2019) Dynamic ion pair behavior stabilizes single α-helices in proteins. J Biol Chem 294(9):3219–3234

Bencharit S, Border MB (2012) Where are we in the world of proteomics and bioinformatics? Expert Rev Proteomics 9(5):489–491

Brown CW, Sridhara V, Boutz DR, Person MD, Marcotte EM, Barrick JE, Wilke CO (2017) Large-scale analysis of post-translational modifications in E. coli under glucose-limiting conditions. BMC Genom 18(1):301

Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Dalenberg K, Duarte JM, Dutta S, Feng Z (2019) RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47(D1):D464–D474

Capasso S, Mazzarella L, Zagari A (1991) Deamidation via cyclic imide of asparaginyl peptides: dependence on salts, buffers and organic solvents. Pept Res 4(4):234–238

Christensen DG, Xie X, Basisty N, Byrnes J, McSweeney S, Schilling B, Wolfe J (2019) Post-translational protein acetylation: an elegant mechanism for bacteria to dynamically regulate metabolic functions. Front Microbiol 10:1604

Clarke S (2003) Aging as war between chemical and biochemical processes: protein methylation and the recognition of age-damaged proteins for repair. Ageing Res Rev 2(3):263–285

Cunin R, Glansdorff N, Pierard A, Stalon V (1986) Biosynthesis and metabolism of arginine in bacteria. Microbiol Rev 50(3):314

Desai PN, Shrivastava N, Padh H (2010) Production of heterologous proteins in plants: strategies for optimal expression. Biotechnol Adv 28(4):427–435

Doyle ML (1997) Characterization of binding interactions by isothermal titration calorimetry. Curr Opin Biotech 8(1):31–35

Englander SW, Mayne L (2014) The nature of protein folding pathways. Proc Natl Acad Sci U S A 111(45):15873–15880

Erlandsen H, Stevens RC (1999) The structural basis of phenylketonuria. Mol Genet Metab 68 (2):103–125

Feke DL, Prabhu ND, Mann JA Jr, Mann JA III (1984) A formulation of the short-range repulsion between spherical colloidal particles. J Phys Chem 88(23):5735–5739

Georgalis Y, Schüler J, Frank J, Soumpasis MD, Saenger W (1995) Protein crystallization screening through scattering techniques. Adv Colloid Interfac 58(1):57–86

Gromiha MM (2010) Protein bioinformatics: from sequence to function. Academic Press; Elsevier, Amsterdam

Hao P, Adav SS, Gallart-Palau X, Sze SK (2017) Recent advances in mass spectrometric analysis of protein deamidation. Mass Spectrom Rev 36(6):677–692

Hu CD, Kerppola TK (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. Nat Biotechnol 21(5):539–545

Hubner IA, Deeds EJ, Shakhnovich EI (2006) Understanding ensemble protein folding at atomic detail. Proc Natl Acad Sci U S A 103(47):17747–17752

Hushcha TO, Luik AI, Naboka YN (2000) Conformation changes of albumin in its interaction with physiologically active compounds as studied by quasi-elastic light scattering spectroscopy and ultrasonic method. Talanta 53(1):29–34

Janson JC (2012) Protein purification: principles, high resolution methods, and applications, vol 151. John Wiley & Sons, Hoboken, NJ

Jaramillo-Flores ME, Soriano-García M, Moreno A (1998) The influence of polyethyleneglycols on predicting crystallisation conditions of lipase from wheat germ by dynamic light scattering studies. J Mol Struct 444(1–3):155–164

Jia L, Sun Y (2017) Protein asparagine deamidation prediction based on structures with machine learning methods. PLoS One 12(7):1–10

Johnson CM (2013) Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys 531(1–2):100–109

Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev 101 (8):2209–2242

Kadima W, McPherson A, Dunn MF, Jurnak F (1991) Precrystallization aggregation of insulin by dynamic light scattering and comparison with canavalin. J Cryst Growth 110(1–2):188–194

Kentache T, Jouenne T, De E, Hardouin J (2016) Proteomic characterization of Nα-and N-ε-acetylation in Acinetobacter baumannii. J Proteome 144:148–158

Kobe B, Jennings IG, House CM, Michell BJ, Goodwill KE, Santarsiero BD, Stevens RC, Cotton RG, Kemp BE (1999) Structural basis of autoregulation of phenylalanine hydroxylase. Nat Struct Biol 6(5):442–448

Laue TM, Stafford WF III (1999) Modern applications of analytical ultracentrifugation. Annu Rev Biophys Biomol 28(1):75–100

Li B, Gorman EM, Moore KD, Williams T, Schowen RL, Topp EM, Borchardt RT (2005) Effects of acidic N+ 1 residues on asparagine deamidation rates in solution and in the solid state. J Pharm Sci 94(3):666–675

Lyumkis D (2019) Challenges and opportunities in cryo-EM single-particle analysis. J Biol Chem 294(13):5181–5197

Matsuura Y, Takehira M, Joti Y, Ogasahara K, Tanaka T, Ono N, Kunishima N, Yutani K (2015) Thermodynamics of protein denaturation at temperatures over 100 C: CutA1 mutant proteins substituted with hydrophobic and charged residues. Sci Rep 5:15545

Mikol V, Vincendon P, Eriani G, Hirsch E, Giege R (1991) Diagnostic of protein crystallization by dynamic light scattering; an application to an aminoacyl-tRNA synthetase. J Cryst Growth 110 (1–2):195–200

Miles AJ, Wallace BA (2016) Circular dichroism spectroscopy of membrane proteins. Chem Soc Rev 45(18):4859–4872

Nachiappan M, Jain V, Sharma A, Manickam Y, Jeyakanthan J (2019) Conformational changes in glutaminyl-tRNA synthetases upon binding of the substrates and analogs using molecular docking and molecular dynamics approaches. J Biomol Struct Dyn 38(6):1575–1589

Nachiappan M, Jain V, Sharma A, Yogavel M, Jeyakanthan J (2018) Structural and functional analysis of Glutaminyl-tRNA synthetase (TtGlnRS) from Thermus thermophilus HB8 and its complexes. Int J Biol Macromol 120:1379–1386

Nikolaev DM, Shtyrov AA, Panov MS, Jamal A, Chakchir OB, Kochemirovsky VA, Olivucci M, Ryazantsev MN (2018) A comparative study of modern homology modeling algorithms for rhodopsin structure prediction. ACS Omega 3(7):7555–7566

Ouidir T, Jarnier F, Cosette P, Jouenne T, Hardouin J (2015) Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14. J Proteome 114:214–225

Parker MW (2003) Protein structure from X-ray diffraction. J Biol Phys 29(4):341–362

Pierce MM, Raman CS, Nall BT (1999) Isothermal titration calorimetry of protein–protein interactions. Methods 19(2):213–221

Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscape theory part I: basic concepts. Q Rev Biophys 35(2):111–167

Reece JB, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB (2011) The formation and function of molecules depend on chemical bonding between atoms. In: Campbell biology, vol 10. Pearson, San Francisco, CA, p 38

Roe S (2001) Protein purification techniques: a practical approach, vol 244. OUP, Oxford

Roth CM, Neal BL, Lenhoff AM (1996) Van der Waals interactions involving proteins. Biophys J 70(2):977–987

Shehu A, Barbara D, Molloy K (2016) A survey of computational methods for protein function prediction. In: Wong KC (ed) Big data analytics in genomics. Springer, Cham

Smyth MS, Martin JHJ (2000) X-ray crystallography. J Clin Pathol 53:8–14

Su XD, Zhang H, Terwilliger TC, Liljas A, Xiao J, Dong Y (2015) Protein crystallography from the perspective of technology developments. Crystallogr Rev 21(1–2):122–153

Sugiki T, Kobayashi N, Fujiwara T (2017) Modern technologies of solution nuclear magnetic resonance spectroscopy for three-dimensional structure determination of proteins open avenue for life scientists. Comput Struct Biotechnol J 15:328–339

Tan YZ, Aiyer S, Mietzsch M, Hull JA, McKenna R, Grieger J, Samulski RJ, Baker TS, Agbandje-McKenna M, Lyumkis D (2018) Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. Nat Commun 9(1):1–11

Tanford C, Reynolds J (2001) Nature's robots: a history of proteins. Oxford University Press, Oxford

Trevino SR, Scholtz JM, Pace CN (2008) Measuring and increasing protein solubility. J Pharm Sci 97(10):4155–4166

Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91

Venkatesan A, Gopal J, Candavelou M, Gollapalli S, Karthikeyan K (2013) Computational approach for protein structure prediction. Healthc Inform Res 19(2):137–147

Walsh MK, Marlon JR, Goring SJ, Brown KJ, Gavin DG (2015) A regional perspective on Holocene fire–climate–human interactions in the Pacific Northwest of North America. Ann Assoc Am Geogr 105(6):1135–1157

Zhang W, Xiao S, Ahn DU (2014) Protein oxidation: basic principles and implications for meat quality. Crit Rev Food Sci 53(11):1191–1201

Wider G (2005) NMR techniques used with very large biological macromolecules in solution. Method Enzymol 394:382–398

Withers P (2013) Landing spacecraft on Mars and other planets: an opportunity to apply introductory physics. Am J Phys 81(8):565–569

Zhen J, Kim J, Zhou Y, Gaidamauskas E, Subramanian S, Feng P (2018) Antibody characterization using novel ERLIC-MS/MS-based peptide mapping. MAbs 10(7):951–959

Zhou HX, Pang X (2018) Electrostatic interactions in protein structure, folding, binding, and condensation. Chem Rev 118(4):1691–1741

Wet-Lab Approaches to Determine Three-Dimensional Structures of Proteins

3

Rajan Kumar Pandey, Rupal Ojha, and Vijay Kumar Prajapati

Abstract

Protein is the functional unit of cells composed of amino acid residues joined together by peptide bond. A total of four forms of proteins are present, starting with the simple primary structure to the complex tertiary and quaternary forms. Tertiary structure is the functional form of protein composed of a single polypeptide chain, while quaternary structure consists of two polypeptide chains. To study the functional aspects of protein, it is necessary to determine its tertiary structure. With the advances of science, various approaches have been reported to produce the tertiary structure either in crystal form or in native stage of proteins. This chapter covers the three scientific techniques, namely nuclear magnetic resonance, X-ray crystallography, and cryo-electron microscopy. These characterization techniques help understand the structure and function of proteins.

Keywords

Protein structure · Nuclear magnetic resonance spectroscopy · X-ray crystallography · Cryo-electron microscopy

3.1 Introduction

Proteins are the complex biological macromolecules consist of a repetitive subunit of amino acids joined together by peptide bond. They have proven to be diverse in terms of functionality, i.e., they are the crucial component of cell membrane providing structural support, helps in the transport of substances, and acts as enzymes (Alberts et al. 2002). Amino acids join together with peptide bonds to

R. K. Pandey · R. Ojha · V. K. Prajapati (✉)
Department of Biochemistry, Central University of Rajasthan, Ajmer, Rajasthan, India
e-mail: vkprajapati@curaj.ac.in

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_3

form a polypeptide chain and one or more polypeptide chains twisting up to form a functional protein. The functional aspect of protein entirely depends on its shape and confirmation. The shape and conformation of a protein depend upon the four levels of protein structures ranging from simple primary structure to the more complex quaternary structure. A simple linear chain of amino acid residues constitutes the primary structure of the protein. These residues joined with each other via a peptide bond, resulting in a polypeptide chain. The next level of protein structure is a secondary structure formed by the interaction between atoms of the backbone of the polypeptide chain that leads to the local folding. The protein secondary structure has two common forms, namely alpha (α) helix and beta (β) sheet, formed by the local folding of the polypeptide chain. These two structural forms have a crucial role in maintaining the shape and conformation of protein molecules by forming hydrogen binding (Alberts et al. 2002). The functional form of protein is a tertiary structure having a three-dimensional structure with more complex protein folding. It forms due to the interaction between R-groups of amino acid residues forming a protein. However, the types of interactions are of non-covalent type, mainly ionic bond, hydrogen bond, London-dispersion forces, and dipole–dipole interaction. The presence of nonpolar and hydrophobic amino leads to their clustering inside of the protein while hydrophilic amino acids occupy an outside position to interact with a water molecule (Godbey 2014). Other interactions include a disulfide bond, a type of covalent bond, formed between sulfur-containing cysteine residues. The proteins are having only one polypeptide chain from the three-level of structure, i.e., primary, secondary, and tertiary. However, proteins having multiple polypeptide chains lead to the formation a quaternary structure. In this case, each polypeptide is known as a subunit (Godbey 2014). The native folded state of the protein has only relatively limited stability over the unfolded state(s) under physiological conditions (Tripathi 2013).

3.2 Approaches to Determine Protein 3D Structure

As above, we discussed the 3D structure of the protein in a defined form, but the actual question is how the 3D structure of a specific protein looks like? Because once we have a protein 3D structure, we can quickly determine the protein functionality, shape, and specificity of the active site. Several approaches are available to determine the protein structure, and each of them has its advantages and disadvantages. Nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, and cryo-electron microscopy (cryo-EM) are the widely used methods to elucidate the structural conformation of the protein (Venien-Bryan et al. 2017) (Fig. 3.1). Each method needs several pieces of information and came out with different experimental data output. X-ray crystallography gives an X-ray diffraction pattern; NMR gives information regarding local conformation of the protein and also uncovers the distance between closely situated atoms. While cryo-EM gives the image elaboration, the overall shape of the molecule. The details of all these methods elucidating protein structure are mentioned as follows.

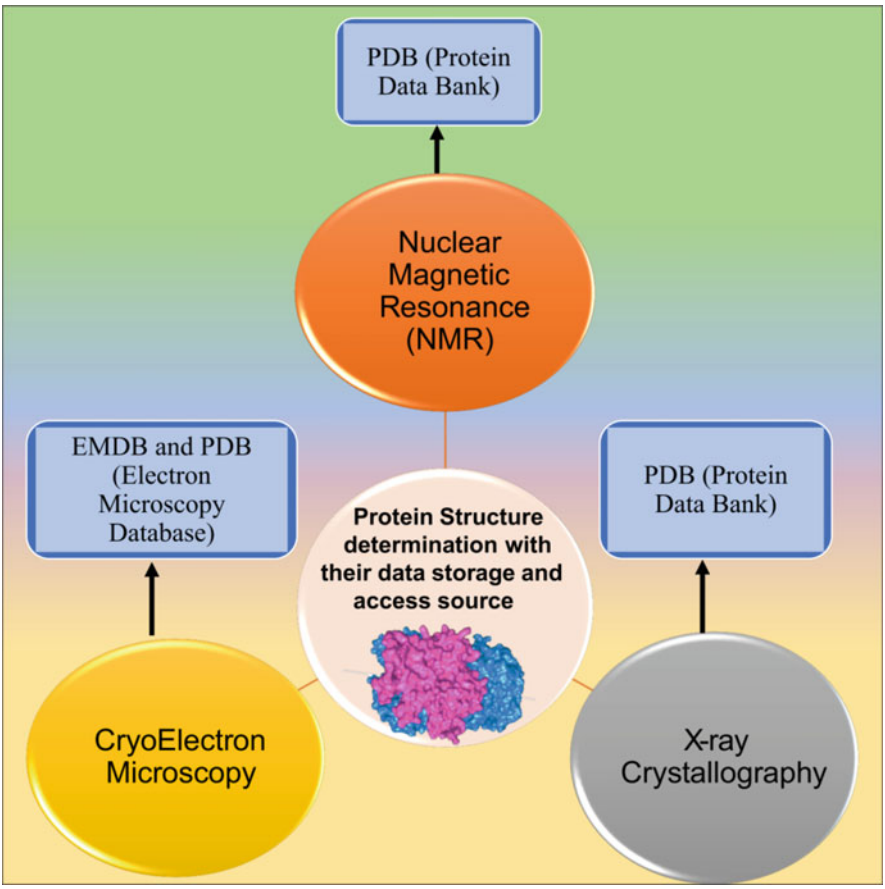


Fig. 3.1 Scientific approaches to determine the protein structure and the available online repositories to access the same

3.2.1 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is an important technology capable of determining the macromolecular structure of the protein molecule in the solution phase and of studying the enzyme kinetics reaction and protein properties at the atomic level. This unique methodology is used to determine the three-dimensional molecular structure of the protein by observing the isotopic signals with a magnetic moment. The isotopes primarily used for the same are ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{32}P, and ^{113}Cd (Kainosho et al. 2006). Cadmium isotopes are rarely used to substitute the protein-bound metal, mainly calcium and zinc. NMR needs a very less quantity of dissolved biological samples (0.2–0.5 mL) with a concentration of 0.2 mM or more. These samples are paced in a glass tube centered in a high magnetic field. Once these samples got irradiated with radiofrequency pulses (also called pulse

sequence), resulting in the generation of the response signal, which reflects the relatedness between different nuclei. These signals can be interpreted to uncover the data about the 3D structure and dynamics state of proteins. Whether the protein is having a sharp and well-characterized 3D structure or having a flexible polypeptide chain, every morphological form exhibits a unique NMR signal. This technique uses a series of protocols to determine the folded protein structure, but its success depends upon the enrichment of protein with stable isotopes. If not improved optimally, then ^{1}H is the only hope to determine the same as this NMR-active constant isotope has a rich profusion in polypeptides. This restricts the utilization of NMR for protein structure determination with a molecular weight of less than 15 kDa. However, the utilization of ^{13}C, ^{15}N, and ^{2}H (rarely) for protein labeling using a bacterial expression system facilitates the structural analysis of protein with a molecular weight ranging 40 kDa or more.

3.2.1.1 Sample Preparation for NMR-Based Structure Determination

Before proceeding NMR-based structure determination, it is essential to prepare the protein sample up to the mark where it can be used for the structure determination. However, it is a time-consuming step. An ideal NMR sample should have a buffer (pH < 8) volume of 0.2–0.5 mL consisting of 0.2–1.0 mM of protein sample. The buffer utilized for the NMR must be steady at room temperature for a more drawn out length. Depending upon the amino acid composition, prepared protein may be insoluble or unstable, but most often, its dissolvability and steadiness can be improved by veering off the salt concentration and change in pH, temperature, and buffer.

Be that as it may, regularly, and dependability can be improved by salt focus and change in pH, temperature, and cradle. This issue can likewise be settled by including a small amount of detergent. Sometimes, soluble enrichment tag (SET) can be used to substitute the surface exposed, hydrophobic, non-essential amino acid residues with the hydrophilic one. This leads to giving the overall solubility of the protein and inhibits the irreversible precipitation (Wüthrich 1986). NMR needs isotopically enrich protein, which can be developed by labeling the protein by using different expression systems, mainly mammalian cell lines, yeast, slime mold, and insect cells. It is the expression vector which decides the amount of protein to be produced and its solubility behavior. The purification of soluble protein is facilitated by adding various tags like six-histidine tag, glutathione *S* transferase (GST-tag), or protein GB1 (GB1) to the expressed protein. These tags must be removed from the expressed protein by enzymatic approaches to get the actual protein before going through the NMR study. The final prepared sample should not consist of more than 5% solute, and it must be devoid of endogenous proteases.

3.2.1.2 Sequence-Specific Labeling of Unlabeled Protein

The process of 1H resonance assignments depends upon the protein size. If the protein size is small, specifically short peptides and small protein, there is no need for isotope labeling for getting the same. In this case, two-dimensional experimental approaches are used. Firstly, each amino acid residue is assigned with the spin

system using the bond ^{1}H–^{1}H correlation with COrrelationSpectroscopY (COSY)-type experiment (Cavanagh et al. 1995). Every one of these experiments may connect over scalar coupling spins isolated by close to three chemical bonds. Further advances deal with the finding of successive through-space networks between the amide protons of peptide spine and protons of preceding amino acid residues. This step comprises a weak interaction, namely dipole–dipole bonding, which gives a cross-peaks in two-dimensional nuclear overhauster effect spectroscopy (NOESY). In conclusion, the successively associated amino spin system is coordinated to the protein sequence. This point is essential and chooses the destiny of the experimentation; here, the spin-system topologies are checked for uniformity.

3.2.1.3 Assigning Progressive Backbone Resonance After Protein Labeling

The protein backbone marking is an essential step toward its structure determination. The most commonly used labeling method is ^{13}C and ^{15}N-based protein labeling, which is helpful to determine protein structure. To achieve an efficient ^{1}H, ^{13}C, and ^{15}N resonance assignment for the protein backbone and side chain, a set of triple resonance ($^{1}H/^{13}C/^{15}N$) experiment can be used (Van de Ven 1995). This arrangement of triple resonance examinations comprises of three experimental pairs, namely HNCACB and HN(CO)CACB, HNCA, and HN(CO)CA, and HNCO and HN(CA)CO. Each letter means that nuclei associated with a specific pulse sequence. The experimental result yields are logged in duos where one amino acid residue based on its N-H group is linked with any one of the signals, namely CA, CO, or the CB signals of own residues or the preceding one. Furthermore, the intra-residue relationship is contrasted with the sequential inter-residue correlation, and the coordinating sets are exposed to the string of spin systems relegated to the protein amino acid residues. A lot of six triple resonance tests can be kept running for a complete length of 6–12 days. The time of the experiment varies as per the protein concentration and molecular weight, but this time can be shortened to 1 day by using cryogenic probes and nonuniform sampling technique (Hoch and Stern 1996). Furthermore, the spectra data can be processed and analyzed by using a computer-based automation program to get the sequential assignment.

3.2.1.4 Side-Chain Assignment for the Labeled Protein

The side chain of labeled proteins can be assigned with the HCCH-TOCSY experiment; leads to fix all ^{13}C and ^{1}H resonance corresponding to the identical amino acid spin system (Cavanagh et al. 1995). Further C(CO)NH and H(CCO)NH experiments are used to correlate aliphatic ^{13}C and ^{1}H resonance with backbone NH. This method is beneficial for correlating all side-chain ^{13}C and ^{1}H resonance with the backbone amide group in a solitary dataset. Along these lines, the task methodology can be effectively mechanized.

3.2.1.5 Resolving Initial Fold

The inter-nuclei distance information can be extracted by getting 2D NOESY and 3D heteronuclear NOESY data. Further, CYANA, CNS, and XPLOR can be used to

determine the initial protein 3D fold with a per residue assigned distance constraint of 3–4 NOE. The computation technique incorporates distance geometrical calculation that utilizes NEO constraints and the simulated annealing method (Craik 1996). At this stage, the separation imperatives identified with the amide of the protein backbone, aromatic side-chain groups, and methyl groups are utmost weighty for characterizing the normal optional structural components and the hydrophobic center of the polypeptide chain. We can use different tools like CANDID and ARIA, to allocate the NEO peak spontaneously and to calculate the early protein fold.

3.2.1.6 Protein Structural Refinement and Validation

It is the early protein model that decides the number of NOE peaks to be assigned. If a higher number of NOE peaks are assigned, it will bring about more separation limitations during the protein refinement. In this repetitive approach, it is furthermore precarious to dispose of incorrect NOE assignments that lead to cause misshaped protein structures and anticipate the assembly of structural scheming. To obtain a good NMR structure, on average, 10–20 constraints per residue is required (James et al. 2001). Several factors may lead to improve the precision and convergence of the NMR-based protein structure ensemble. These factors are backbone carbon dihedral angle constrain, hydrogen bond constraints, side-chain rotamers libraries, orientation constraints, and stereospecific assignment of methyl or methylene group.

The final ensemble NMR structure should be routinely checked for its quality, and this could be done by using computational software, namely ProcheckNMR. There are specific rules to asses and validate the NMR structural output quality, mainly (a) the usual RMSD of the backbone atom should be less than 1.0 in comparison with mean structure, (b) the number of residues should not be more than 1% in the disallowed region, once validated for the Ramachandran plot (c) there should not be more than 0.5 of NOE distance violation. The dihedral angular violation should be less than 5°.

3.2.2 X-Ray Crystallography

X-ray diffraction-based sub-atomic structure analysis has gained its importance in the early twentieth century. In the late nineteenth century, X-ray was discovered by Wilhelm Conrad Röntgen, and he received the first Nobel Prize for Physics for this discovery. Later by 1912, X-ray diffraction in the crystal was discovered by Friedrich, Von Laue, and Knipping. However, Bragg and Brag were the pioneers to obtain the crystal structure; for that, they received Nobel Prize for physics in 1915s. X-ray crystallography can be aimed to elucidate the three-dimensional structural feature of the biological macromolecules from a crystal. The macromolecules may be either protein or nucleic acid but must be purified and concentrated on making their crystal followed by exposure to the X-ray beam. It is one of the most broadly utilized and incredible tools to elucidate the

three-dimensional form of biological macromolecules with the atomic level of resolution. The data output of the X-ray diffraction pattern can be used to determine the crystal packaging symmetry and the size of the atoms that form the protein crystal (Picknett and Brenner 2001). The strengths of diffraction spots can be utilized to explain the structural elements to map the density of electron. This step followed by the quality improvement of the map to authorize the making of protein molecular structure using its primary polypeptide sequence. The resulting structure is then refined to opt for a thermodynamically favored conformation. The whole steps of X-ray crystallography, starting from crystallization to model building, are summarized as follows.

3.2.2.1 Protein Crystallization

The protein crystallography starts with the expression, purification, and concentration of protein to yield high-quality, homogenous, and soluble protein. The obtained protein is used for crystallization purposes, but the growing protein crystal must be of sufficient quality because it is the factor that decides the fate of protein crystallographic work. The protein crystallization starts with the induction of highly concentrated protein to come out of a solution under standard condition because fastening of this process leads to the protein precipitation (Luft et al. 1994; Wilson et al. 1991; Gernert et al. 1988). This is the step that determines the success of a proposed work because the inability of protein crystallization may abort the whole project. The success of protein crystallization can be assured by optimizing different variables, including protein concentration, temperature, pH, buffer, and the crystallization technique. Commercially available crystal screen packages can be used to optimize these variables and yield the protein crystals of sufficient quality (Jancarik and Kim 1991; Hampel et al. 1968). Generally, the glycosylated, flexible, and less conformationally constrained proteins are tough to crystalize while proteins with large complexity but high symmetry are easy to crystalize (Rossmann et al. 1985; Smyth et al. 1995). The ideal protein crystal should have a longest dimension of minimum 0.1 mm to offer a sufficient crystal lattice volume that can be subjected to the X-ray beam.

3.2.2.2 Optical Setup for X-Rays

The generation of an optimum X-rays beam is an essential step of crystallography post protein crystallization. There are two ways to generate an X-ray, mainly accelerated electrons in a synchrotron storage ring or by using an electron striking a copper anode. The former follows the process of monochromatic, where we select the only single X-ray wavelength while remaining others were excluded by absorption, while in the later no need of the same because it generates a strong wavelength. Among any of them, the main objective is to focus the ray into the beam and afterward collimated to guarantee the parallel nature of the beam. Further, X-ray beams are made accurately parallel with adjustable slits to 0.1–0.3 mm diameter. Then the protein crystal is subjected to the beam of X-ray, and the goniometer is wisely adjusted to ensure its presence in the beam.

3.2.2.3 Diffraction Analysis

Post-exposure of macromolecule crystals to the X-ray beam, the next step is to analyze the diffraction data. X-ray exposure to protein crystal can be given via two modes, firstly by X-ray generator or by synchrotron sources. The latter is advantageous over the first one in generating a powerful beam of an X-ray with high-quality optics; this results in need of shorter exposure time but comes out with a higher signal-to-noise ratio among the diffraction image. This method is favorably used to get the solution of challenging crystallographic problems (Helliwell 2005; Ealick and Walter 1993). The crystal can be exposed to the X-ray beam by using two methods mainly by exposing the protein crystal after placed into the capillary tube at normal room temperature or place it in liquid nitrogen using a small size loop (Hampel et al. 1968). The latter uses a single protein crystal but gives a complete dataset. Another side the former has the advantage of getting diffraction characteristics of a new protein crystal. Initially, X-ray films were used to collect the X-ray diffraction images, but in the past two decades, imaging plates took its place because it is highly sensitive (approx. ten times or more) than conventional X-ray film and read out a digitized image within a few minutes of completion of the exposure. However, charged coupled device (CCD) has replaced the imaging plate by their readout times in seconds rather than minutes (Moy 1994; Gruner 1994). Before X-ray exposure, it is good to calculate and justify the distance between the protein crystal and the detector. This will lead to collect the diffracted spots as a rule with a limit of 1.5–3.0 Å goals. The resolution of the collected spot is directly proportional to the diffracted angle on the detector.

Like this, the essential goals will be acquired at the edge of the detector, and if we choose the diffracted edge required and the separation between the detector and crystal protein, it very well may be effectively balanced. To get the principal diffraction picture for another crystal, it must affirm that the diffraction prompts high resolution to make a protein structure with clear details at near-atomic level. For the most part, a spot resolution up to 3 Å is adequate to give detailed for amino acid side chains in the electron density map.

3.2.2.4 Data Collection

It is the condition which decides the quantity of data needed and plan for its achievement (Dauter et al. 1997). For example, (1) a cubic crystal system having high symmetry only needs to collect the X-ray diffraction data over as small as 35°. On the other side, a crystal with lower symmetry needs to collect the diffraction data over 180°. (2) For non-crystallographic symmetry such as virus particles having may, identical subunits constitute the highest level of NCS (Smyth et al. 1995). Here, the high-quality structure can be gotten from informational indexes that are a long way from 100% complete in light of remuneration by averaging. Other factors include (3) the requirement of molecular replacement, and (4) the upper-resolution limit requirement. Initial diffraction investigation comes out with the indication of spots positioning on the detector at an ideal distance. This data can be further processed to decide the oscillation (range for each X-ray exposure). During X-ray beam exposure, crystal is rotated perpendicular to the beam of ray, and this leads to

the recording of a maximum number of data for each image. It is the size of crystal that decodes the angle of diffraction, for example, the larger unit has an angle of approximately 0.25°, and for the smaller unit cells, an angle of 2° is much enough to collect spot without overlap.

3.2.2.5 Processing of Output Data

Softwares-based algorithms have made easier the mathematical processing of output diffraction data using well-established algorithms. These algorithms are easy to understand, and even newcomers can easily process the data and calculate the resulting electron density map. The data processing can be divided into various steps (1). The first step includes resolving the crystal system and the dimension of the unit cell in an accurate manner. This step also allows determining all the promising orientation for the crystal in the beam of X-ray (Gruner 1994). The process of indexing can only be processed once we have knowledge of the cell and its positioning (Kabsch 1988). There are three numerals, namely h, k, and l, which are used to assign an index to each spot of the image. This process should be thoroughly checked to avoid possible errors due to mis-indexing. The next step consists of the intensity measurement for each spot. In general, it was absorbed that protein crystal diffract weakly due to the presence of light atoms and large unit cells in their composition. However, the larger the crystal volume, the stronger will be the diffraction. The amplitude and phase relation of the diffracted waves decide the intensities of diffracted spots. DENZO is the most important computational program used in protein crystallography to measure the intensity and perform auto-indexing (Otwinowski and Minor 1997). Secondly, there must be a scale factor that can be used to relate the intensities of all the images, and it allocated at a scale factor of one for the first image, and subsequent images will be scaled up to attain the same. Based on the scaling statistics, the picture will be dismissed or reprocessed to save its quality for the benefit of the final dataset. This step can be easily handled by using SCALEPACK program (Otwinowski and Minor 1997). The output scaling file can be sorted by using the SortMTZ program of the CCP4 suite.

3.2.2.6 Amplitude Analysis

The diffracted waves and its incident on the detector decide the intensity of the diffracted spots. The amplitude of diffracted waves and the angle between them called phase differences. Different phase difference values have different interference; constructive interference can be achieved at a lower phase difference of zero, while the higher phase difference value of 180° leads to destructive interference. There are several computational tools, mainly the Truncate module of CCP4, which can be used to calculate the amplitude (Collaborative CP 1994).

3.2.2.7 Electron Density Mapping and Calculation

Once we have the data of amplitude and phases, we can go with the calculation of structural factors by using the fast Fourier transform (FTT) (Ten Eyck 1973). The succeeding electron density map is the critical factor to frame the 3D shape of a protein crystal, which further leads to construct the protein structure. The spacing

present at the edge of the unit cell decides the nature of the map detail and the speed of the figuring. This makes the three-dimensional grid inside the unit cell.

3.2.2.8 Refinement and Model Building

Refinement is an important step in improving the quality of the electron density map (Brünger 1992). If an asymmetric unit consists of multiple identical molecules or subunits, it might be possible to use their molecular average to make the refined structure, and this can be achieved by using the NCS. Further, the observed structural features are compared with the one which we calculated to determine the inconsistency known as R-factor. The phasing power is proportional to the square root of the NCS. Once we have an electron density map that has sufficient quality, then the model is built. Further, computational tools can be used to visualize the map.

3.2.3 Cryo-Electron Microscopy

Cryo-electron microscopy (cryo-EM) is an advanced technique used to visualize the structural features of proteins. The presence of cryo-EM would have been not possible without the contribution of Dubochet and colleagues (1980s) in unique specimen preparation protocol (Adrian et al. 1984) to preserve the biological material in its native state using amorphous ice film of a little thickness. This directly allows observing the molecular structure using a low-dose transmission electron microscope functioning at a very low temperature of liquid nitrogen or even below. This elucidates the structural features of biological macromolecules at near-atomic resolution in comparison with NMR and X-ray crystallography. The major advantage of this method is the lack of necessity of 3D protein crystal as it supports to observe the protein in multiple conformations under its natural environment. This accommodates a more extravagant understanding of the dynamic behavior of input protein molecules. Secondly, cryo-EM can determine the protein structure with a starting molecular weight of around 50 kDa, which is the upper limit of NMR.

In cryo-EM, specimen preparation is an important step where biomolecules are rapidly frozen in solution followed by their loading into an electron microscope column working at very low temperature but having a high vacuum which maintains the amorphous state of ice, thereby minimizing the radiation-mediated damages to the protein (Fujiyoshi 2013). Depending on temperature, ice has two forms, mainly cubic (115–150 K) or hexagonal crystal form (170 K). However, the rapid freezing of biomolecules bypasses the crystal form but makes amorphous ice. Liquid nitrogen is the best way to provide the most reliable data quality; however, it also provides the specimen protection from damage compared with the low temperature.

With the increasing time, electron microscopy has gained two significant updates, mainly the detector and image processor. The former got upgraded by the entry of a direct electron detector (DED), which directly detects the electron and greatly enhanced the quantum efficiency as compared to previous detectors (Brilot et al. 2012). The latter is the image processing algorithm, which constantly increased the microprocessor performance and allowed classifying various EM images.

3.2.3.1 Sample Preparation for Cryo-EM

This is the first and most important step of cryo-EM starting with sample verification to the low-dose image collection. A solution containing macromolecules is applied to a holey carbon EM grid. These grids consist of various conductive materials, ranging from copper to gold and many others. These grids further get covered by a perforated support layer having a regular array of holes of well-defined size, shape, and pitch. The hydrated sample with a volume of 2–3 mL is applied to the holey grid, followed by removing the excessive solution, and placing this grid into cryogen already precooled using liquid nitrogen. One thing must be kept in mind to distribute the sample by pre-treatment of the support film evenly. For the soluble sample, the hydrophilic surface is required, which is created by bombing ions in a glow-discharging device or by using plasma cleaning (Bernecky et al. 2016). However, if the sample is hydrophobic, then amyl amine can be used in the evaporation chamber (Grassucci et al. 2007). The most common cryogen is ethane-propane mix having the advantage of reducing freezing temperature to allow the direct thermal contact among cryogen and liquid nitrogen devoid of freezing the mixture. Presently, blotting robots like FEI Vitrobot, Gatan CP3, or Leica EM GP can be used for ice embedding of specimens to remove excess liquid. This process needs very less volume of moderate concentration protein (0.1–2.0 mg/mL). Currently, multiple approaches are under research and development to improve cryoembedding.

3.2.3.2 Low-Dose Image Collection

Low-dose procedures are helpful to yield high-resolution images of cryo-EM for the ice-embedded biological macromolecules on a digital image detector. This method is also helpful in avoiding irradiation damage to the biological macromolecules by defocused diffraction mode. The frozen grid containing biomolecules is placed on cryo-transfer holder, and then this specimen is laden into cryo-EM by the help of cryo-transfer holder. Cryo-specimen cassettes can also be used instead of the cryo-transfer holders and having the advantage of holding a dozen of cartridge-mounted grids simultaneously. Initially, the focus of cryo-EM is adjusted nearby, then, the target is given with exposure at specific preset magnification to get clear images.

3.2.3.3 Single-Particle Analysis for Cryo-EM

Single-particle analysis (SPA) is a powerful technique used to determine the protein structures that have escaped crystallization. Principally, this is an average view of multiple copies of the primary molecule, therefore, sometimes called single-particle averaging. The SPA information accumulation technique disseminates the doses of electron among the multiple copies of alike biomolecules. Since every molecule has arbitrary orientation once frozen, there is no need to tilt the specimen containing stage to achieve complete rotational sampling.

Moreover, regular SPA datasets include a great many pictures, all of which contain up to a thousand projections of the biomolecules in different directions. The image reconstruction consists of two alternative steps: (1) search for possible orientation for all projections and (2) 3D reconstruction, which leads to making a three-dimensional electron potential map. This cycle is typically iterated until the

resolution of the obtained picture does not improve further. To achieve increased phase contrast, the defocused condition is applied to collect cryo-EM images. These pictures are modified in frequency space by the phase contrast transfer function (CTF). This can be assessed by using various computational softwares, for example, ctffind, gctf, and e2ctf.py (Rohou and Grigorieff 2015; Zhang 2016; Bell et al. 2016). Post CTF estimation, randomly oriented molecular images are poised, and various tools can be used to automatically pick the particle and make their correlation. These tools are scipion, signature, FindEM, Relion-autopick, and e2boxer.py (de la Rosa-Trevín et al. 2016; Chen and Grigorieff 2007; Roseman 2004; Scheres 2012; Tang et al. 2007). Lastly, the 3D structure is reconstituted for the biomolecules in its native state.

3.2.3.4 Subtomogram Averaging

Subtomogram averaging (STA) is a recent success of tomograms, and this is the basis of STA. The tomograms of closely identical biomolecule 3D structure are the starting point of STA. The procedure starts with information gathering by tilt series pursued by tomographic remaking. Further, during following up the averaging method, the main emphasis should be given to the first frame because of its least damage due to radiation effect. The reconstructed tomogram came out with the selection of multiple copies of particles to build subtomograms or 3D particles. STA prompts the expanded SNR of the final subtomogram average and simultaneously reducing the missing wedge artifacts. Further, CTF corrections can be useful to each projection data before image reconstruction.

3.2.3.5 Resolution Analysis, Model Building, and Validation

The comparison of two reconstructions leads to the estimation of the possible resolution of reconstruction. The two reconstructions come from half a set of images in the frequency domain, and they differ more and more with increasing levels of details. This correlation coefficient can be plotted in the form of a Fourier shell correlation (FSC). The threshold FSC of 0.5 indicates the resolution of the map at minimized reference bias.

Lastly, all approaches above come out with the structural features and dynamics of biomolecule, mainly protein. Once we obtain a high-resolution structure of biomolecule reconstructed beyond 4 Å, it is frequently promising to assemble a nuclear model of the whole molecule de novo or by using rigid-body docking. This is followed by the automated structure refinement by incorporating chemical knowledge and force field. Further, the refined model must be subjected to the validation using a crystallographic geometry validation tool. The output map deposited into the public database, namely Electron Microscopy Data Bank and Protein Data Bank.

References

Adrian M, Dubochet J, Lepault J, McDowall AW (1984) Cryo-electron microscopy of viruses. Nature 308:32–36

Alberts B, Johnson A, Lewis J (2002) Molecular biology of the cell. Garland Science, New York
Bell JM, Chen M, Baldwin PR, Ludtke SJ (2016) High resolution single particle refinement in EMAN2.1. Methods 100:25–34
Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P (2016) Structure of transcribing mammalian RNA polymerase II. Nature 529:551–554
Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N (2012) Beam-induced motion of vitrified specimen on holey carbon film. J Struct Biol 177:630–637
Brünger AT (1992) X-PLOR: version 3.1: a system for x-ray crystallography and NMR. Yale University Press, New Haven, CT
Cavanagh J, Fairbrother WJ, Palmer AG, Skelton NJ (1995) Protein NMR spectroscopy: principles and practice. Elsevier, Burlington, MA
Chen JZ, Grigorieff N (2007) SIGNATURE: a single-particle selection system for molecular electron microscopy. J Struct Biol 157:168–173
Collaborative CP (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr Sect D: Biol Crystallogr 50:760–763
Craik DJ (1996) NMR in drug design. CRC Press, Boca Raton
Dauter Z, Lamzin VS, Wilson KS (1997) The benefits of atomic resolution. Curr Opin Struct Biol 7:681–688
de la Rosa-Trevín JM, Quintana A, del Cano L, Zaldívar A, Foche I, Gutiérrez J et al (2016) Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. J Struct Biol 195:93–99
Ealick SE, Walter RL (1993) Synchrotron beamlines for macromolecular crystallography. Curr Opin Struct Biol 3:725–736
Fujiyoshi Y (2013) Low dose techniques and cryo-electron microscopy. In: Schmidt-Krey I, Cheng Y (eds) Electron crystallography of soluble and membrane proteins: methods and protocols. Humana Press, Totowa, NJ, pp 103–118
Gernert KM, Smith R, Carter DC (1988) A simple apparatus for controlling nucleation and size in protein crystal growth. Anal Biochem 168:141–147
Godbey WT (2014) Chapter 2—Proteins. In: Godbey WT (ed) An introduction to biotechnology. Woodhead Publishing, Cambridge, pp 9–33
Grassucci RA, Taylor DJ, Frank J (2007) Preparation of macromolecular complexes for cryo-electron microscopy. Nat Protoc 2:3239–3246
Gruner SM (1994) X-ray detectors for macromolecular crystallography. Curr Opin Struct Biol 4:765–769
Hampel A, Labanauskas M, Connors PG, Kirkegard L, Rajbhandary UL, Sigler PB, Bock RM (1968) Single crystals of transfer RNA from formylmethionine and phenylalanine transfer RNA's. Science 162:1384–1387
Helliwell JR (2005) Macromolecular crystallography with synchrotron radiation. Cambridge University Press, Cambridge
Hoch J, Stern A (1996) NMR data processing. Wiley-Liss. Inc., New York, NY
James TL, Dotsch V, Schmitz U (2001) Nuclear magnetic resonance of biological macromolecules. Methods Enzymol 339:238–258
Jancarik J, Kim SH (1991) Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Crystallogr 24:409–411
Kabsch W (1988) Automatic indexing of rotation diffraction patterns. J Appl Crystallogr 21:67–72
Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Ono AM, Güntert P (2006) Optimal isotope labelling for NMR protein structure determinations. Nature 440:52–57
Luft JR, Arakali SV, Kirisits MJ, Kalenik J, Wawrzak I, Cody V, Pangborn WA, Detitta GTA (1994) Macromolecular crystallization procedure employing diffusion cells of varying depths as reservoirs to tailor the time course of equilibration in hanging-and sitting-drop vapor-diffusion and microdialysis experiments. J Appl Crystallogr 27:443–452

Moy JP (1994) A 200 mm input field, 5–80 keV detector based on an X-ray image intensifier and CCD camera. Nucl Intrum Meth A 348:641–644
Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326
Picknett T, Brenner S (2001) X-ray crystallography. In: Encyclopedia of genetics. Elsevier, Amsterdam
Rohou A, Grigorieff N (2015) CTFFIND4: fast and accurate defocus estimation from electron micrographs. J Struct Biol 192:216–221
Roseman AM (2004) FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. J Struct Biol 145:91–99
Rossmann MG, Arnold E, Erickson JW, Frankenberger EA, Griffith JP, Hecht HJ et al (1985) Structure of a human common cold virus and functional relationship to other picornaviruses. Nature 317:145–153
Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519–530
Smyth M, Tate J, Hoey E, Lyons C, Martin S, Stuart D (1995) Implications for viral uncoating from the structure of bovine enterovirus. Nat Struct Biol 2:224–231
Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, Ludtke SJ (2007) EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 157:38–46
Ten Eyck LF (1973) Crystallographic fast Fourier transforms. Acta Crystallogr Sect A 29:183–191
Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91
Van de Ven FJ (1995) Multidimensional NMR in liquids: basic principles and experimental methods. VCH, New York, Weinheim, Cambridge
Venien-Bryan C, Li Z, Vuillard L, Boutin JA (2017) Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. Acta Crystallogr Sect F 73:174–183
Wilson L, Bray T, Suddath F (1991) Crystallization of proteins by dynamic control of evaporation. J Cryst Growth 110:142–147
Wüthrich K (1986) NMR with proteins and nucleic acids. Europhys News 17:11–13
Zhang K (2016) Gctf: real-time CTF determination and correction. J Struct Biol 193:1–12

4 Use of Group-Specific Reagents in Active Site Functional Group Elucidation I: Cys, Ser, Tyr, and Trp Residues

Pravin Kumar Ambasht

Abstract

The functional groups present in the active site of an enzyme are explored with the help of the determination of pK_a values and group-specific reagents. Some of the commonly used group-specific reagents are listed along with the functional groups: iodoacetamide, iodoacetate, *p*-chloromercuribenzoate, *N*-ethylmaleimide (Cys); phenylmethylsulfonyl fluoride (Ser); tetranitromethane, iodine (Tyr); *N*-bromosuccinimide (Trp). The enzyme is incubated with group-specific reagent and its effect is tested on activity. The loss of activity indicates the presence of the functional group. The modification is also analyzed through UV-visible, fluorescence, and circular dichroism spectra analysis. The inactivation kinetics, kinetic parameters, protection in the presence of substrates, etc. have been elaborated.

Keywords

N-ethylmaleimide · Iodoacetamide · *N*-bromosuccinimide · Enzyme inactivation · Activity · Kinetics

4.1 Introduction

Enzymes are biological catalysts that are responsible for catalyzing reactions in the living system. They are basically proteins; however, there are reports of catalytic RNA (ribozyme) and catalytic antibodies (abzymes). In the present chapter, I am going to consider enzymes that are proteins. Enzymes have a cleft responsible for the binding of substrate and bringing about its transformation to products. The

P. K. Ambasht (✉)
Department of Biochemistry, School of Life Sciences, North-Eastern Hill University, Shillong, Meghalaya, India
e-mail: pambasht@nehu.ac.in

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_4

functional groups in the active site are, therefore, important for a proper understanding of the catalytic process. Proteins have different levels of structures like primary, secondary, tertiary, and quaternary. The primary structure represents the covalent linear structure with an amino acid sequence without any interaction. In secondary structure, there is folding represented by an alpha helix and a beta-pleated sheet. In both, there is the formation of hydrogen bonds between carbonyl oxygen (C=O) and amino hydrogen (–NH) of the peptide bond. The tertiary structure involves superfolding, giving rise to a compact native structure in the case of non-oligomers. The hydrogen bonds are between side chains. Other interaction includes salt bridges and van der Waals force. The quaternary structure includes oligomeric enzymes with the interaction between subunits. Under physiological conditions, the folded state of the protein has very limited stability over the unfolded state (Tripathi 2013).

It is observed that in the native state, the polar residues are exposed and interact with water. The nonpolar residues are, however, buried and protect themselves from water. A few polar residues are also buried that form the active site cleft. The active site residues in primary structure may lie far apart, but in the native state, they lie in close proximity. Chymotrypsin has Ser195, His57, and Asp102 constituting the active site (Stryer 1995). The active site was considered to have a rigid structure. Now, it is known to be flexible with the induced fit model proposed by Koshland. The latter suggested that with the approach of the substrate toward the enzyme, the shape of active site changes. In the case, of carboxypeptidase A, there is a movement of Arg145 and Glu270 by 2 Å and Tyr248 by 12 Å upon binding of substrate. The substrate is supposed to be in close proximity and correctly oriented for interaction with functional groups of the active site. The substrate binds to the active site and forms the enzyme–substrate complex and the interaction is close to less than 0.4 nm.

To study the mechanism of action of an enzyme, it is necessary to investigate the functional groups in the active site. The use of group-specific reagents was developed to investigate amino acid residues present in the active site. In addition, the use of affinity labeling was also employed. Further, site-directed mutagenesis was used as an important tool in this direction. In the present chapter, the focus is on the use of group-specific reagents. The effects of group-specific reagents on enzyme activity, kinetics, and spectral analysis have been presented. In the present chapter, the focus is on the investigation of Cys, Ser, Tyr, and Trp residues in the enzyme active site. In the subsequent chapter, the discussion is focused on Glu, Asp, Arg, Lys, and His residues.

4.2 Cysteine

The thiol group is one of the important functional groups present in the active site of an enzyme. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is one such enzyme that catalyzes the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate in the presence of NAD^+ and Pi (Stryer 1995). Cysteine proteases depend on the thiol group for catalytic functions (Lowe 1976). Thiol group

functions as a nucleophile and, therefore, can be modified with greater ease, and the reactions have been reviewed earlier (Liu 1977; Brocklehurst 1979).

4.2.1 Reversible and Irreversible Blockers

The thiol functional group can be modified in the presence of reversible and irreversible blockers. The reversible blockers include fluorodinitrobenzene (FDNB), sodium tetrathionate, 4,4′-dithiodipyridine, and 5,5′-dithiobis-(2-nitrobenzoic acid) (DTNB). The latter two are helpful in quantitative determination (Kenyon and Bruice 1977). Quantification of thiols and disulfides has been explained and reviewed (Yan et al. 1998; Winther and Thorpe 2014). DTNB and 4,4′-dithiodipyridine react with –SH group and liberate 2-nitro-5-thiobenzoic acid (Ellman 1959) and 4-dithiopyridone. 2-Nitro-5-thiobenzoic acid absorbs strongly at 410 nm (Ellman 1959). In some reports, absorbance was recorded at 412 nm (Song et al. 1986). Reagents used for estimation of –SH group in a protein and the ones keeping the reducing environment have been shown in Fig. 4.1.

The enzyme is titrated with excess DTNB in the absence and presence of sodium dodecyl sulfate (SDS). In the absence of SDS, the result indicates reactive SH groups while, in the presence of SDS, the result indicates total –SH groups (Kumar and Malhotra 1988; Malhotra and Kayastha 1989). In place of SDS, 8 M urea has also been used (Malhotra et al. 1991).

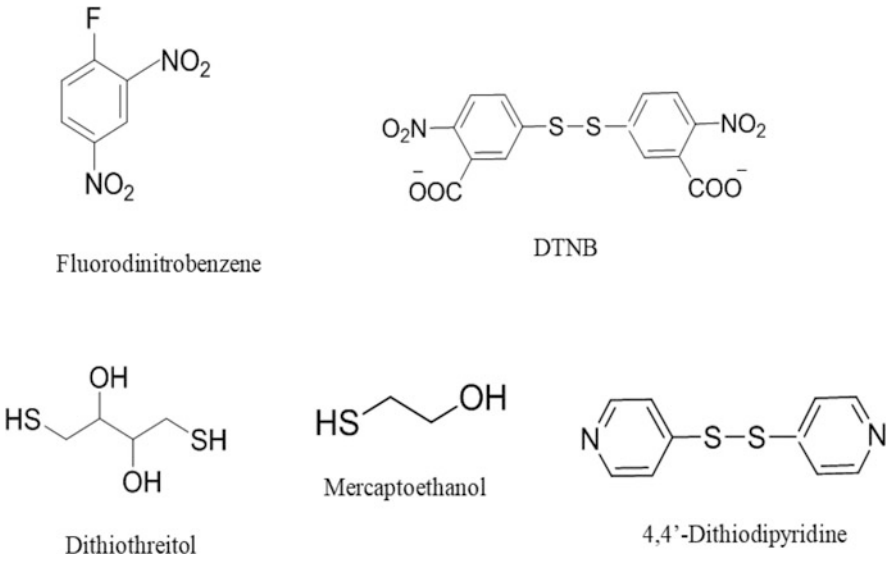


Fig. 4.1 Structures of different thiol reagents employed in the estimation of –SH group in a protein and under reducing environment

4.2.2 Estimation of Thiol

PEP phosphatase from mung bean is a tetrameric enzyme with a total of 8 –SH groups. Four –SH groups were found to be more reactive. The reaction of DTNB with the enzyme was carried out at pH 7.5 and 8.5, exhibited a biphasic curve. There is a distinct fast and slow phase. The amplitude of fast and slow phases was nearly similar at both pH; however, values of k_{fast} and k_{slow} were higher at pH 8.5. The overall result suggested that two of the reactive –SH groups react faster than the other two reactive –SH groups (Malhotra and Kayastha 1989). The reports on the use of DTNB for the estimation of thiol groups have been summarized in Table 4.1.

4.2.3 Thiol Modifiers

The thiol functional group present in the active site can be modified with the following reagents which block the SH group irreversibly, namely *p*-chloromercuribenzoate (*p*-CMB), *N*-ethylmaleimide (NEM), iodoacetamide, iodoacetate, and bromoacetate (Kenyon and Bruice 1977). Bromoacetate can also be used, but the reaction is slower in comparison to iodoacetate as iodide is a better leaving group (Eyzaguirre 1987).

Dissociation of catalytic and regulatory subunits of rabbit muscle cAMP-dependent protein kinase has been shown to occur in the presence of *p*-CMB (Murray et al. 1974). The structures of thiol group-specific reagents and structure of modified active site thiol have been shown in Figs. 4.2 and 4.3. Molecules like dithiothreitol and β-mercaptoethanol have been used in extraction buffers for keeping a reducing environment (Malhotra and Kayastha 1990; Ambasht et al. 1996).

Table 4.1 Use of DTNB for the estimation of thiol groups

S. no.	Enzyme	Reactive thiol	Total thiol	References
1.	Sulfhydryl oxidase	1.93	3.29	Song et al. (1986)
2.	Phosphoglycerate kinase	1.03	3.3	Kumar and Malhotra (1988)
3.	Phospho*enol*pyruvate carboxylase	7.85	12.1	Stiborova (1988)
4.	PEP-phosphatase	3.88	8.07	Malhotra and Kayastha (1989)
5.	Isocitrate lyase	4.15	7.9	Malhotra et al. (1991)
6.	Urease	5.82	12.1	Srivastava and Kayastha (2000)
7.		6	34	Kumar and Kayastha (2010)

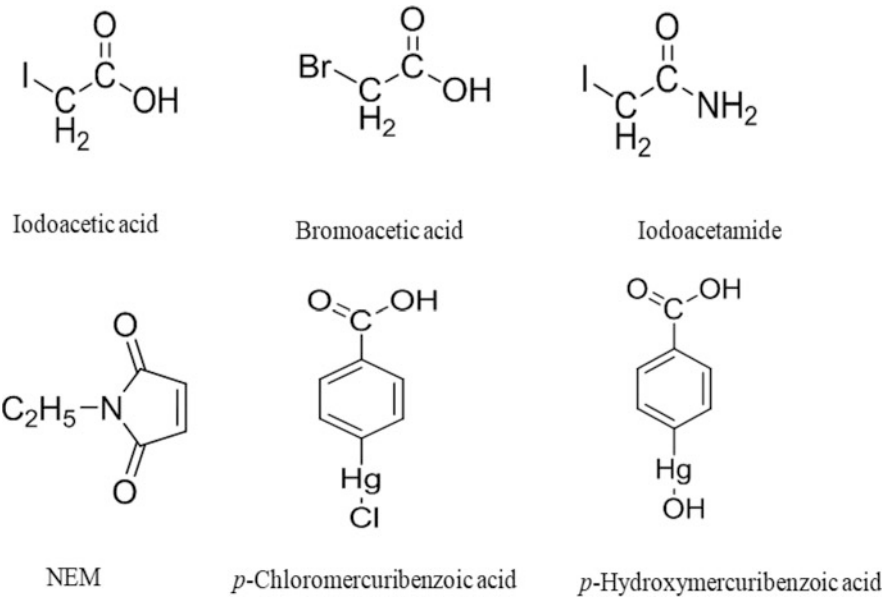


Fig. 4.2 Thiol group-specific reagents employed to modify thiol groups in an enzyme

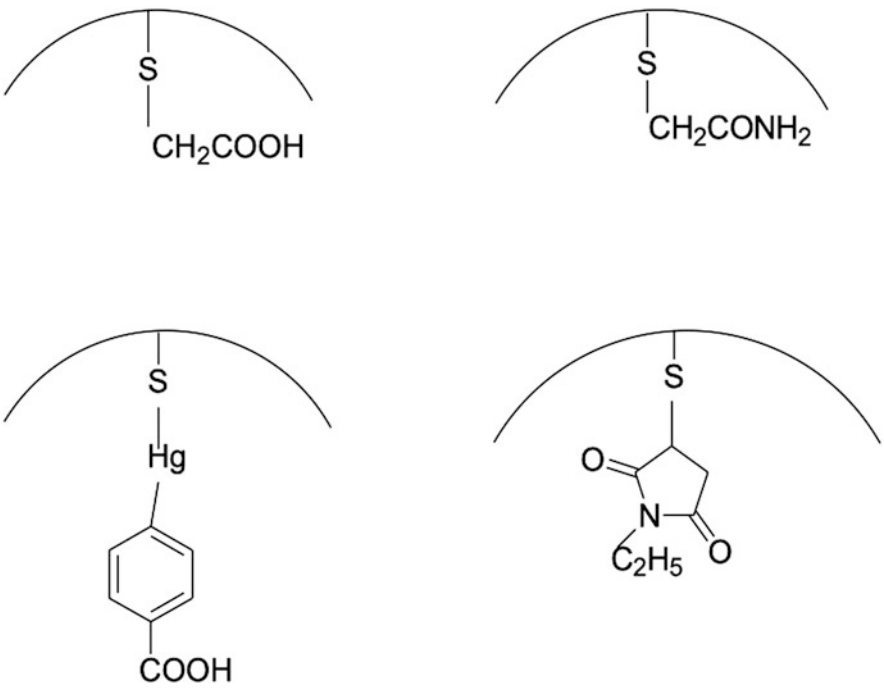


Fig. 4.3 Modified thiol group in the presence of thiol group-specific reagents

4.2.4 Reaction

Iodoacetamide and iodoacetate are known as alkylating agents and react with sulfhydryl groups under neutral pH (Suzuki 2015). Iodoacetamide reacts with the free thiol group of a protein covalently forming stable carboxamide methylated protein and HI. NEM and *p*-CMB react with –SH group under acidic conditions and form adduct and mercaptide, respectively (Riordan and Vallee 1972a). NEM exhibits a reaction between pH 5.0 and 7.0, while *p*-CMB between pH 4.5 and 5.0 (Suzuki 2015). NEM is a hydrophilic molecule, while *p*-CMB and DTNB are having hydrophobic structural elements (Malhotra et al. 1991). The reaction of NEM with the thiol group shows better specificity and can be monitored spectrophotometrically. The organic mercurial *p*-CMB dissolves in water and becomes hydroxyl derivative and reacts with thiol, displaying an increase in absorbance at 255 nm (Eyzaguirre 1987). The reaction of the thiol group in the presence of thiol reagent exhibits an increase in absorbance at 255 nm due to the formation of a complex between Cys residue and the reagent (Huang and Ichikawa 1995; Roknabadi et al. 1999). The use of different thiol reagents on different enzymes available in literature has been summarized in Table 4.2. NEM and *p*-CMB have been used mostly in comparison to alkylating agents. Thiol-specific reagents like NEM, iodoacetamide, and *p*-CMB have been used for the identification of cysteine proteases (Kundu et al. 2000; Duarte et al. 2009; Singh et al. 2010).

4.2.5 Inhibition Studies

Cytosolic 5′-nucleotidase and plasma membrane 5′-nucleotidase exhibited contrasting results when exposed to iodoacetamide (15 mM) for 15 min at 37 °C. The plasma membrane enzyme was very sensitive, with only 4.0% residual activity, while the cytosolic enzyme showed a marginal loss in activity (92% residual activity). The plasma membrane enzyme displayed protection in the presence of ATP. In the presence of $HgCl_2$, both enzymes showed inactivation. In the presence of NEM, both enzymes displayed marginal losses in the activity. The overall result suggests the presence of a thiol group in the active site (Worku et al. 1984). Maize phospho*enol*pyruvate carboxylase activity was lost when it was pre-incubated with *p*-CMB (147 μM) and Cu^{2+}(5 μM) for 30 min. In both cases, PEP (1 mM) exhibited protection in the activity. The enzyme showed activation in the presence of DTT and β-mercaptoethanol. The result suggests the involvement of the thiol group in the catalytic function of the enzyme (Stiborova 1988). NEM brought inactivation of 11β-hydroxysteroid dehydrogenase ($IC_{50} = 10$ μM). The activity was enhanced by DTT ($EC_{50} = 1$ mM) (Niu and Yang 2002).

Invertase from *Trapaeolum* was competitively inhibited in the presence of *p*-CMB. The thiol group present in the active site was protected by fructose but not by sucrose and trehalose. The enzyme was also inhibited in the presence of DTNB (Isla et al. 1998). GAPDH from rabbit muscle exhibited a loss in activity in the presence of peroxynitrite (Souza and Radi 1998).

Table 4.2 Use of different thiol reagents

S. no.	Thiol reagents	Enzyme	Reference
1.	Iodoacetamide	Pig heart malate dehydrogenase	Aspray et al. (1979)
2.		Plasma membrane 5′-nucleotidase	Worku et al. (1984)
3.		Cytosolic 5′-nucleotidase	Worku et al. (1984)
4.		*Klebsiella aerogens* urease	Todd and Hausinger (1991)
5.		*Corynebacterium* sarcosine oxidase	Hayashi et al. (1983), Suzuki and Kawamura-Konishi (1991)
6.		Mung bean pyruvate kinase	Ambasht et al. (1997)
7.		Pigeon pea urease	Srivastava and Kayastha (2000)
8.		Platelet α-1,6-fucosyl transferase	Kaminska et al. (2003)
9.		Soybean urease	Kumar and Kayastha (2010)
10.	Iodoacetate	Sulfhydryl oxidase	Song et al. (1986)
11.		Mung bean phosphoglycerate kinase	Kumar and Malhotra (1988)
12.		Mung bean PEP-phosphatase	Malhotra and Kayastha (1989)
13.		Mung bean GAPDH	Malhotra et al. (1993)
14.		Watermelon seed urease	Prakash and Bhushan (1998)
15.		Platelet α-1,6-fucosyl transferase	Kaminska et al. (2003)
16.		Castor bean acid phosphatase	Granjeiro et al. (2003)
17.		*Jacaratia* protease	Duarte et al. (2009)
18.		Porcerain B	Singh et al. (2010)
19.	NEM	Urease	Gorin and Chin (1965)
20.		Jack bean urease	Andrews and Reithel (1970)
21.		Plasma membrane 5′-nucleotidase	Worku et al. (1984)
22.		Cytosolic 5′-nucleotidase	Worku et al. (1984)
23.		Mung bean phosphoglycerate kinase	Kumar and Malhotra (1988)
24.		Mung bean PEP-phosphatase	Malhotra and Kayastha (1989)
25.		*Brassica nigra* PEP-phosphatase	Duff et al. (1989)
26.		Chalcone isomerase	Bednar (1990)
27.		Castor isocitrate lyase	Malhotra et al. (1991)
28.		Mung bean GAPDH	Malhotra et al. (1993)
29.		Rat liver $NADP^+$ linked isocitrate dehydrogenase	Fatania et al. (1993)
30.		Mung bean pyruvate kinase	Ambasht et al. (1997)
31.		α3-Fucosyl transferase	Britten and Bird (1997)
32.		Watermelon seed urease	Prakash and Bhushan (1998)
33.		Pigeon pea urease	Srivastava and Kayastha (2000)
34.		Ervatamin B	Kundu et al. (2000)

(continued)

Table 4.2 (continued)

S. no.	Thiol reagents	Enzyme	Reference
35.		11β-hydroxysteroid dehydrogenase	Niu and Yang (2002)
36.		Platelet α-1,6-fucosyl transferase	Kaminska et al. (2003)
37.		Porcerain B	Singh et al. (2010)
38.		Soybean urease	Kumar and Kayastha (2010)
39.	*p*-CMB	Phospho*enol*pyruvate carboxylase	Stiborova (1988)
40.		Invertase	Isla et al. (1998)
41.		Mung bean phosphoglycerate kinase	Kumar and Malhotra (1988)
42.		Mung bean PEP-phosphatase	Malhotra and Kayastha (1989)
43.		Castor isocitrate lyase	Malhotra et al. (1991)
44.		Retinal oxidase	Huang and Ichikawa (1995)
45.		Mung bean pyruvate kinase	Ambasht et al. (1997)
46.		Watermelon seed urease	Prakash and Bhushan (1998)
47.		Pigeon pea urease	Srivastava and Kayastha (2000)
48.		Ervatamin B	Kundu et al. (2000)
49.		Platelet α-1,6-fucosyl transferase	Kaminska et al. (2003)
50.		Porcerain B	Singh et al. (2010)
51.		Soybean urease	Kumar and Kayastha (2010)
52.	*p*-HMB	Barley β-glucosidase	Skoubas and Georgatos (1997)
53.		Lentil acid phosphatase	Roknabadi et al. (1999)
54.	Peroxynitrite	Rabbit muscle GAPDH	Souza and Radi (1998)

4.2.6 Inactivation Kinetics

The effect of thiol reagents was tested on the activities of different enzymes. Phosphoglycerate kinase from mung bean did not exhibit any loss in activity in the presence of iodoacetate, NEM, and *p*-CMB at 5 mM concentration, suggesting that –SH groups are not critical for the enzyme activity (Kumar and Malhotra 1988).

PEP-phosphatase from *Brassica nigra* also did not exhibit a loss in the presence of NEM (Duff et al. 1989; Ambasht and Kayastha 1999). Barley β-glucosidase did not show inactivation in the presence of *p*-HMB (5 mM), suggesting that Cys is not part of the active site (Skoubas and Georgatos 1997).

Castor bean acid phosphatase displayed inactivation in the presence of iodoacetic acid at 37 °C but not at 25 °C, suggesting that it is temperature-dependent enzyme. The inactivation followed second-order kinetics and exhibited protection against inactivation in the presence of Pi and *p*-nitrophenol, suggesting that thiol is part of the active site (Granjeiro et al. 2003). Single exponential loss in activity was observed in GAPDH in the presence of peroxynitrite with IC_{50} of 17 μM. The activity loss was along with protein thiol oxidation (8 –SH/tetramer); however, only

one critical SH group was important. Glyceraldehyde-3 phosphate exhibited protection (Souza and Radi 1998).

Some of the enzymes, upon reaction with thiol reagents, exhibited pseudo-first-order kinetics (Song et al. 1986; Malhotra et al. 1993; Fatania et al. 1993; Huang and Ichikawa 1995; Ambasht et al. 1997; Britten and Bird 1997; Roknabadi et al. 1999). Interestingly, mung bean GAPDH followed pseudo-first-order kinetics with iodoacetate and NEM, but the biphasic kinetic pattern with *p*-CMB (Malhotra et al. 1993). In the case of $NADP^+$-linked isocitrate dehydrogenase from rat liver, $NADP^+$ brought enhancement of inactivation, while isocitrate brought protection against inactivation (Fatania et al. 1993). From the rate versus pH plot of mung bean pyruvate kinase catalyzed reaction, pK_a value for side chain was determined and found to be 8.9, suggesting the presence of thiol group tentatively at the active site. With iodoacetamide, NEM, and *p*-CMB at 100 μM, inactivation was observed with k values 0.022, 0.032, and 0.029 min^{-1}, respectively (Ambasht et al. 1997). Lentil non-specific acid phosphatase in the presence of 100 μM *p*-HMB exhibited an identical value of k (0.115 min^{-1}) without substrates *p*-NPP and AMP (Roknabadi et al. 1999). In fucosyl 3-transferase, NEM brought 50% inhibition at 200 μM and complete inhibition at 500 μM. The inactivation at 500 μM is time- and temperature-dependent. GDP-fucose displayed protection against NEM inhibition (Britten and Bird 1997).

Some other enzymes exhibited biphasic inactivation kinetics in the presence of thiol reagents with distinct slow and fast phases (Malhotra and Kayastha 1989; Malhotra et al. 1991; Prakash and Bhushan 1998; Srivastava and Kayastha 2000; Kumar and Kayastha 2010). It is being suggested that biphasic kinetic pattern and half-site reactivity are observed with large nonpolar groups (Malhotra et al. 1993). The time course of inactivation is consistent with the following equation:

$$A_t = A_{\text{fast}}{}^{e-k\text{fast}.t} + A_{\text{slow}}{}^{e-k\text{slow}.t}$$

where A_t = % residual activity at time t, A_{fast} and A_{slow} are the amplitudes, and k_{fast} and k_{slow} are first-order rate constants.

In most of the enzymes, A_{fast} and A_{slow} were equal to 50% or close to it (Malhotra and Kayastha 1989; Malhotra et al. 1993; Srivastava and Kayastha 2000; Kumar and Kayastha 2010). In the case of watermelon urease, A_{fast} and A_{slow} were 60 and 40%, respectively, in the presence of NEM (Prakash and Bhushan 1998).

Protection against inactivation (*p*-CMB) was also observed in the presence of L-Cys (Malhotra et al. 1993; Srivastava and Kayastha 2000; Kumar and Kayastha 2010). Fluoride exhibited protection against NEM inactivation (Srivastava and Kayastha 2000; Kumar and Kayastha 2010). The substrate urea provided much weaker protection against SH group reagents (Srivastava and Kayastha 2000).

Native jack bean urease exhibited a fast reaction in the presence of DTNB and NEM with 26–28 SH groups without inactivation, but further titration brought reaction with 7–9 –SH groups slowly with inactivation (Andrews and Reithel 1970). Urease reacts with NEM (21 moles) rapidly without any loss in activity, but further reaction with NEM (7–8 moles) was slow with loss in activity (Gorin and

Chin 1965). Urease from *Klebsiella aerogenes* exhibited inactivation in the presence of [^{14}C] iodoacetamide at pH 6.3 with the incorporation of radioactivity into a single peptide, and protection was observed in the presence of Pi. When a similar reaction was carried out at pH 8.5, the modification was observed in several peptides. However, modification of a single peptide was identical to that observed at pH 6.3 in the loss of activity pattern. A single thiol group (Cys-319) was identified to be essential for catalysis (Todd and Hausinger 1991).

Malate dehydrogenase from pig heart was alkylated in the presence of iodoacetamide, giving rise to carboxamidomethyl cysteine and carboxamidomethyl methionine. The results suggest that the inactivation of the enzyme was not due to the above modifications (Aspray et al. 1979). The effect of iodoacetamide on sarcosine oxidase from *Corynebacterium* was investigated and was found to inhibit the activity. In the presence of sodium acetate, the inhibition was prevented. Sodium sulfite did not bring any inhibition to the iodoacetamide-treated enzyme in the presence or absence of sodium acetate. Results suggest that there are at least two Cys residues that are located in the subunit B, one at the sarcosine-binding site and other to the covalent FAD-binding site (Suzuki and Kawamura-Konishi 1991). In the presence of NEM, chalcone isomerase got inactivated following pseudo-first-order kinetics below pH 7.0. At pH 9.9, the semi-log plot was not linear and the curvature is due to base-dependent decomposition of NEM. The inactivation was dependent on the concentration of NEM. Over a range of pH, the results suggest modification of a single Cys residue. Competitive inhibitors protected the activity, suggesting that this Cys residue is near the active site (Bednar 1990). Sarcosine oxidase from *Corynebacterium* in the presence of iodoacetamide was inactivated with complete loss in activity. The inactivation followed first-order kinetics with respect to iodoacetamide concentration. In the presence of acetate, 50% protection was observed (Hayashi et al. 1983). α-1,6-Fucosyltransferase from human platelets was subjected to different thiol reagents like *p*-CMB, iodoacetamide, iodoacetate, NEM, marasalyl, and methyl iodide. All reagents except NEM brought inactivation. GDP-fucose and GDP offered protection against inactivation. The result suggests that Cys residue is close to the substrate binding site (Kaminska et al. 2003).

4.3 Serine

Amino acid serine possesses –OH group in its side chain. Its role in the mechanism of action of chymotrypsin has been well elaborated (Stryer 1995). Different reagents are being used to block –OH group of a Ser residue in an enzyme. Alkyl isocyanates have been found to inhibit Ser proteases. Enzymes like chymotrypsin, elastase, and trypsin displayed different behavior toward alkyl isocyanate. Trypsin activity is unaffected by octyl and butyl isocyanate; elastase is inactivated by only butyl isocyanate, whereas chymotrypsin is inhibited by both the above inhibitors (Brown and Wold 1973). Phenylmethylsulfonyl fluoride (PMSF) is one of the blockers of a Ser residue. Barley β-glucosidase, when incubated with PMSF, did not bring any change in activity, and thus, the possibility of serine to be part of active

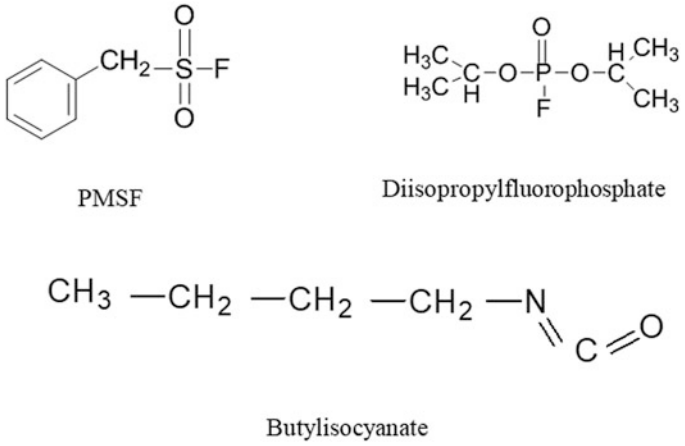


Fig. 4.4 Serine group-specific reagents

site was ruled out (Skoubas and Georgatos 1997). Group-specific reagents are shown in Fig. 4.4.

Different enzymes were shown to exhibit inactivation in the presence of PMSF (Tavakoli et al. 2006; Maruthiah et al. 2013; Matkawala et al. 2019). Choline oxidase from *Alcaligenes* was modified with PMSF, and the Lineweaver-Burk plot indicated uncompetitive inhibition. Further, the fluorescence spectrum revealed that there was no change in the intrinsic fluorescence in the presence of PMSF. The far UV-CD results exhibited no significant change upon modification. The overall results suggested the location of Ser outside the active site (Tavakoli et al. 2006). Some of the proteases displayed inactivation in the presence of PMSF and activation in the presence of mercaptoethanol and DTT (Maruthiah et al. 2013; Matkawala et al. 2019). Protease from *Bacillus subtilis* AP-MSU-6 displayed complete loss in activity at 10 mM PMSF and thus establishing it to be a serine protease (Maruthiah et al. 2013). Protease from *Neucosmospora* is suggested to be a thiol-dependent serine protease (Matkawala et al. 2019). PMSF has been used for the identification of several serine proteases (Pande et al. 2006; Yadav et al. 2006; Singh et al. 2008; Kumari et al. 2010; Raskovic et al. 2014). Diisopropyl fluorophosphate has also been reported to be used for the identification of serine proteases (Pande et al. 2006; Yadav et al. 2006). Diisopropylfluorophosphate specifically reacts with Ser-195 only of chymotrypsin but not with other 27 Ser residues (Stryer 1995). The structure of modified Ser residue in the presence of PMSF has been shown in Fig. 4.5.

4.4 Tyrosine

Tyrosine is an aromatic amino acid with the characteristic phenolic group at the para-position in the side chain. Its presence in the active site is being detected through several compounds that bring about its modification. Commonly used ones are

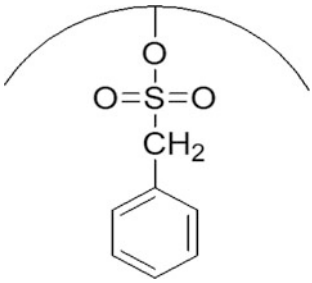


Fig. 4.5 Modified Ser residue in the presence of PMSF

tetranitromethane (TNM) and iodine, which bring about nitration and iodination, respectively. The use of tetranitromethane has been elaborated for nitration of Tyr residues (Sokolovsky et al. 1966; Sokolovsky and Riordan 1970; Riordan and Vallee 1972b). The mechanism of nitration in the aromatic compounds and Tyr has been reviewed (Isaacs and Abed 1982; Ischiripoulos 2009).

4.4.1 Methods, Reaction, and Characteristics

TNM is dissolved in 95% alcohol, while iodine is dissolved in KI (Krishnan and Dua 1985). TNM reacts with Tyr under the mild condition to yield chromophore 3-nitrotyrosine, which has a p*K* value close to 7.0 and is more acidic than Tyr (Eyzaguirre 1987; Suzuki 2015). The reaction of Tyr with iodine upon modification yields diiodotyrosine (Ramchandran 1956).

TNM is also known to bring about dimerization or trimerization of the enzyme (Riordan and Vallee 1972b). The cross-linking of protein molecule was ruled out in the case of modified α-amylase and glutamine synthetase as there was no change in the molecular weight (Kochhar and Dua 1985a; Krishnan and Dua 1985). Further, the modification of aminopeptidase M did not bring about dissociation of subunits as the elution volume of the native, and the modified enzyme was identical (Femfert and Pfleiderer 1969). The structure of Tyr modifying reagents and modified Tyr residue has been shown in Figs. 4.6 and 4.7, respectively.

4.4.2 Reports on the Identification of Tyr Residues and Stoichiometry

Reports on the identification of Tyr residues in the active site are summarized in Table 4.3. The number of Tyr residues modified are summarized in Table 4.4.

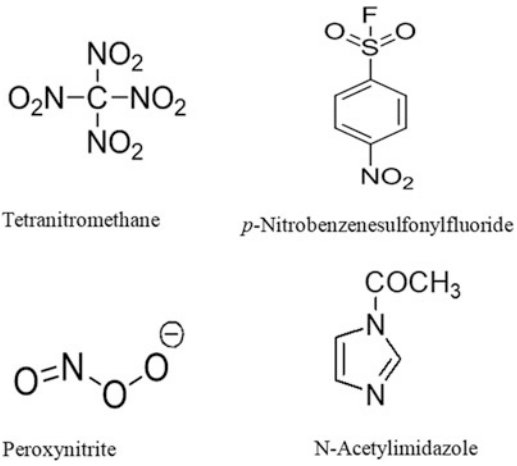


Fig. 4.6 Group-specific reagents for Tyr residue

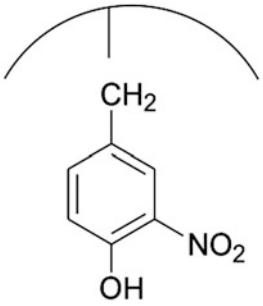


Fig. 4.7 Modified Tyr residue in the presence of TNM

4.4.3 Spectral Analysis

TNM is also known to modify Cys and Trp residues with characteristic absorption spectrum between 300 and 400 nm (Riordan and Vallee 1972b; Kochhar and Dua 1985a). The modification of Trp and Cys was ruled out in α-amylase as there was no absorption peak between 300 and 400 nm (Kochhar and Dua 1985a). The nitrotyrosyl content in the TNM-modified enzymes was detected by recording absorption at 428 and 430 nm (Femfert and Pfleiderer 1969; Maralihalli and Bhagwat 1992; Yang et al. 1996). Modification of Tyr as acetyl-tyrosine in α-amylase has been reported in the presence of acetylimidazole, and the difference absorption spectrum of native and modified enzyme exhibited a decrease at 280 nm (Kochhar and Dua 1985a).

4.4.4 Inactivation Kinetics

The effect of TNM and iodine has been tested on mung bean enzymes like phosphoglycerate kinase, PEP-phosphatase, and pyruvate kinase. These enzymes did not exhibit any loss in activity, thus ruling out the presence of the Tyr group in the active

Table 4.3 Reagents used for the identification of Tyr residues

S. no.	Reagent	Enzyme	Reference
1.	TNM	Aminopeptidase M	Femfert and Pfleiderer (1969)
2.		Glutamine synthetase	Krishnan and Dua (1985)
3.		α-Amylase	Kochhar and Dua (1985a)
4.		α-Galactosidase	Mathew and Balasubramaniam (1986)
5.		Phosphoglycerate kinase	Kumar and Malhotra (1988)
6.		PEP-phosphatase	Malhotra and Kayastha (1989)
7.		Phospho*enol*pyruvate carboxylase	Maralihalli and Bhagwat (1992)
8.		Uridine phosphorylase	Komissarov and Debabov (1995)
9.		H^+ pyrophosphatase	Yang et al. (1996)
10.		Transketolase	Kovina et al. (1996)
11.		Polygalacturonase	Stratilova et al. (1996)
12.		Pyruvate kinase	Ambasht et al. (1997)
13.		Glutathione-*S*-transferase	Lee and Fung (2003)
14.	Iodine	Glutamine synthetase	Krishnan and Dua (1985)
15.		Phosphoglycerate kinase	Kumar and Malhotra (1988)
16.		PEP-phosphatase	Malhotra and Kayastha (1989)
17.		Pyruvate kinase	Ambasht et al. (1997)
18.	Peroxynitrite	Tyrosine hydroxylase	Blanchard-Fillion et al. (2001)
19.		Glycogen phosphorylase	Dairou et al. (2007)
20.	*N*-Acetylimidazole	α-Amylase	Kochhar and Dua (1985a)
21.		Polygalacturonase	Stratilova et al. (1996)
22.		Glycogen phosphorylase	Dairou et al. (2007)
23.	Nitroglycerine	Glutathione-*S*-transferase	Lee and Fung (2003)
24.	Acetic anhydride	Polygalacturonase	Stratilova et al. (1996)
25.	*p*-Nitrobenzene sulfonyl fluoride	Phospholipase A2	Yang et al. (1985)

site (Kumar and Malhotra 1988; Malhotra and Kayastha 1989; Ambasht et al. 1997). Pyruvate kinase exhibited a loss in activity in the presence of DEPC without any change in absorbance at 278 nm, ruling out the presence of Tyr residue in the active site (Ambasht et al. 1997).

Loss in activity was observed when the enzymes were incubated with TNM (Krishnan and Dua 1985; Kochhar and Dua 1985a; Mathew and Balasubramaniam 1986; Maralihalli and Bhagwat 1992; Komissarov and Debabov 1995; Kovina et al. 1996; Stratilova et al. 1996; Yang et al. 1996; Lee and Fung 2003). The inactivation in α-galactosidase from coconut was dependent on TNM concentration. The nitration initially was fast and subsequently slowed down due to the breakdown of reagent. The inactivation was there at pH 8.0 and not below pH 6.0, thus ruling

Table 4.4 Modification of Tyr residues

S. no.	Reagent	Enzyme	Number of modified Tyr	Reference
1.	*p*-Nitrobenzene sulfonyl fluoride	Phospholipase A2	2	Yang et al. (1985)
2.	TNM	α-Amylase	5	Kochhar and Dua (1985a)
3.		α-Galactosidase	1	Mathew and Balasubramaniam (1986)
4.		Uridine phosphorylase	4	Komissarov and Debabov (1995)
5.		Transketolase	3	Kovina et al. (1996)
6.		H^+ pyrophosphatase	4	Yang et al. (1996)
7.	*N*-acetyl imidazole	α-Amylase	5	Kochhar and Dua (1985a)

out the modification of Cys residue (Mathew and Balasubramaniam 1986). In maize PEP carboxylase, the inactivation was studied at pH 6.3 and 7.9, and it was found that the extent of inactivation increased at pH 7.9. Further, they masked the –SH group with p-HMB before modification with TNM (Maralihalli and Bhagwat 1992). In most cases, the reaction of TNM with enzyme has been carried out at pH 8.0 (Kovina et al. 1996; Stratilova et al. 1996). Inactivation results suggested the presence of essential Tyr residue for catalysis (Stratilova et al. 1996; Yang et al. 1996). Bacterial glutamine synthetase and uridine phosphorylase exhibited pseudo-first- and second-order kinetics, respectively, upon treatment with TNM (Krishnan and Dua 1985; Komissarov and Debabov 1995).

Inactivation of enzyme activity was also observed when the enzymes were modified in the presence of peroxynitrite (Blanchard-Fillion et al. 2001; Dairou et al. 2007), *N*-acetylimidazole (Kochhar and Dua 1985a; Stratilova et al. 1996), acetic anhydride (Stratilova et al. 1996), and *p*-nitrobenzene sulfonyl fluoride (NBSF) (Yang et al. 1985). The modification of Tyr residue of glycogen phosphorylase by peroxynitrite brought inactivation following second-order kinetics. Further, the modification of Tyr was confirmed through inactivation in the presence of *N*-acetylimidazole. No loss in activity was observed with NEM and iodoacetamide, thus ruling out the modification of a thiol residue (Dairou et al. 2007). Tyrosine hydroxylase, when exposed to 250 μM peroxynitrite, more than 50% loss in activity was observed without oxidation of thiol. No loss in activity was observed in the mutant enzyme (Tyr-423 → Phe). The results suggest that nitration of Tyr-423 is responsible for the inactivation of tyrosine hydroxylase (Blanchard-Fillion et al. 2001). Tyr-3 and Tyr-63 residues from snake venom phospholipase A2 were modified in the presence of *p*-nitrobenzenesulfonyl fluoride at pH 8.0 with p*K* values 10.1 and 11.0, respectively. The reactivity of Tyr-3 toward NBSF was not affected by the presence or absence of Ca^{2+} ions. However, Tyr-63 exhibited enhanced activity by Ca^{2+} ions (Yang et al. 1985).

The inactivation of polygalacturonase in the presence of *N*-acetylimidazole was concentration-dependent and was fast initially. When *N*-acetylimidazole was 60-fold molar excess, 50% loss was achieved within 60 min. However, when the reagent was 170-fold, molar excess, 80% loss was achieved within 10 min (Stratilova et al. 1996). In α-amylase, when the reagent was 200-fold excess, 40% loss in activity was achieved after 30 min. In the presence of hydroxylamine, full activity was restored (Kochhar and Dua 1985a).

4.4.5 Kinetic Parameters

Modification of uridine phosphorylase in the presence of TNM led to a shift of pH optimum from 7.1–7.2 to 6.5–6.6. In the presence of protective ligands, similar pH optimum as that of native enzyme was exhibited (Komissarov and Debabov 1995). Modified plant vacuolar pyrophosphatase displayed an increase in the K_m without a change in V_{max}. There is a possibility that modification of Tyr residue may induce conformational change, thereby lowering the accessibility of substrate (Yang et al. 1996). Modification of Tyr residue of α-amylase brought an increase in apparent K_m value for soluble starch (Kochhar and Dua 1985a).

4.4.6 Protection Studies

Substrates and products displayed protection against inactivation (Kochhar and Dua 1985a; Krishnan and Dua 1985; Mathew and Balasubramaniam 1986; Maralihalli and Bhagwat 1992; Komissarov and Debabov 1995; Yang et al. 1996). In glutamine synthetase, Glu and ATP together displayed protection suggesting the importance of Tyr residue in transferase and synthetase activities (Krishnan and Dua 1985). Substrate PEP and Mg^{2+} brought complete protection against TNM inactivation in PEP carboxylase from maize (Maralihalli and Bhagwat 1992). Substrate protection studies in liquefying α-amylase from *Bacillus amyloliquefaciens,* suggest that only two Tyr residues were protected and thus suggesting the participation of two Tyr residues in the active site. The reduction of TNM-modified α-amylase with sodium dithionite brought a partial restoration of activity (Kochhar and Dua 1985a). Inactivation of glutathione-*S*-transferase activity was observed when the enzyme was treated with TNM in the absence of GSH. In the presence of the latter, however, protection of activity was observed (Lee and Fung 2003).

4.5 Tryptophan

Tryptophan is an aromatic amino acid with an indole ring, and it plays a vital role in the catalysis of different enzymes as part of an active site. Various compounds have been explored for the identification of tryptophanyl residue in the active site of an enzyme. Most reports are there on the use of *N*-bromosuccinimide (NBS) for the

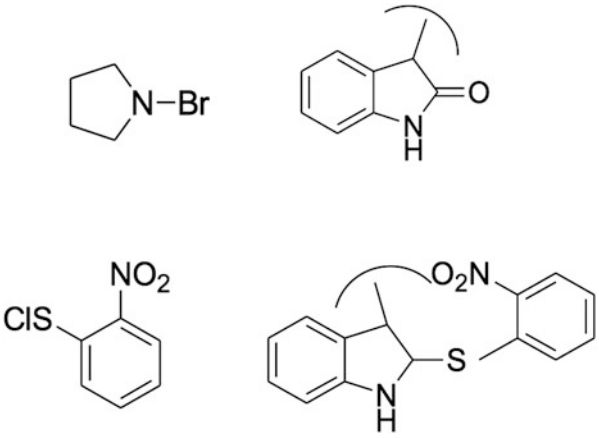


Fig. 4.8 Group-specific reagents for modification of Trp residue and corresponding modifications

modification of Trp residue. NBS is an oxidizing agent and brings about the formation of oxindole (Spande and Witkop 1967; Eyzaguirre 1987). NBS is not very specific for Trp residue and reacts with also Cys, His, and Tyr (Spande and Witkop 1967). The structures of NBS and 2-nitrophenylsulfenyl chloride, along with modifications are shown in Fig. 4.8.

4.5.1 Methods

The number of Trp residues modified is calculated using the molar absorption coefficient ($5.5 \times 10^3\ M^{-1}\ cm^{-1}$) (Spande and Witkop 1967). The NBS-mediated inactivation is followed by monitoring fall in absorption at 280 nm (Gote et al. 2007). The results of the use of NBS and 2-nitrophenylsulfenyl chloride for modification of Trp residues in the active site are summarized in Table 4.5.

4.5.2 Effect of Incubation of NBS on Enzymes

Some enzymes, when incubated with NBS, did not show any inactivation suggesting that Trp is not important for catalysis (Kumar and Malhotra 1988; Malhotra and Kayastha 1989; Ambasht et al. 1997). Some other enzymes exhibited a loss in activity when incubated in the presence of NBS (Kochhar and Dua 1985b; Mathew and Balasubramaniam 1986; Hsieh et al. 1993; Rawat and Rao 1996; Mardanyan et al. 2001; Gote et al. 2007; Kestwal and Bhide 2007). In some cases, the NBS/enzyme ratio was found to be important for modification of Trp (Nagashima

Table 4.5 Use of group-specific reagents for modification of tryptophanyl residue

S. no.	Reagent	Enzyme	Reference
1.	NBS	α-Amylase	Kochhar and Dua (1985b)
2.		Lipoxygenase 1	Klein et al. (1985)
3.		α-Galactosidase	Mathew and Balasubramaniam (1986), Gote et al. (2007)
4.		Aspartate aminotransferase	Nagashima et al. (1986)
5.		Phosphoglycerate kinase	Kumar and Malhotra (1988)
6.		PEP-phosphatase	Malhotra and Kayastha (1989)
7.		Glucokinase	Hsieh et al. (1993)
8.		Xylose reductase	Rawat and Rao (1996)
9.		Pyruvate kinase	Ambasht et al. (1997)
10.		Arginyl t-RNA synthetase	Zhang et al. (1998)
11.		Lipoxygenase-1	Srinivasulu and Appu Rao (2000)
12.		Glutaryl 7-aminocephalo-sporanic acid acylase	Lee et al. (2000)
13.		Adenosine deaminase	Mardanyan et al. (2001)
14.		β-Galactosidase	Kestwal and Bhide (2007)
15.		Amylomaltase	Rachadech et al. (2015)
16.	2-Nitrophenyl-sulfenyl chloride	Phospholipase A2	Chang et al. (1993)

et al. 1986; Srinivasulu and Appu Rao 2000; Mardanyan et al. 2001). At low NBS/enzyme ratio, in aspartate aminotransferase from pig heart cytosol, modification of Trp-122 occurred without any loss in activity. However, at a higher NBS/enzyme ratio, Trp-48 was modified with loss in activity (Nagashima et al. 1986). At a higher NBS/enzyme ratio, a decrease in activity of soybean lipoxygenase and bovine adenosine deaminase from brain and spleen was observed (Srinivasulu and Appu Rao 2000; Mardanyan et al. 2001).

4.5.3 Stoichiometry

The reports on the determination of some modified Trp residues are summarized in Table 4.6. The role of pH is important in the accessibility of Trp residue toward NBS. In soybean lipoxygenase-1, only one Trp residue was accessible for NBS at pH 9.0, and it increased to 7 at pH 2.0 (Srinivasulu and Appu Rao 2000).

4.5.4 Inactivation Kinetics

The loss of activity of the enzyme in the presence of NBS was found to be pH-dependent. α-Galactosidase from *Bacillus stearothermophillus* exhibited an 80% loss in activity at pH 3.0, but no loss was observed at pH 5.0 (Gote et al. 2007). The loss of enzyme activity in the presence of NBS was also concentration-

Table 4.6 Stoichiometry of modified Trp residues

S. no.	Reagent	Enzyme	Number of modified Trp residues	Reference
1.	NBS	Lipoxygenase-1	2	Klein et al. (1985)
2.		α-Galactosidase	1	Mathew and Balasubramaniam (1986)
3.		Arginyl-t-RNA synthetase	1	Zhang et al. (1998)
4.	2-Nitrophenyl-sulfenyl chloride	Phospholipase A2	3	Chang et al. (1993)

dependent (Lee et al. 2000; Gote et al. 2007). Amylomaltase from *Coorynebacterium glutamicum*, when treated with NBS, resulted in inactivation following pseudo-first-order kinetics. Substrate protection indicated that modified Trp residue is part of the active site (Rachadech et al. 2015). In another case, biphasic kinetics in the inactivation was observed with a fast and slow phase and the order was 2 (Rawat and Rao 1996). Sulfenylation of Trp-18 of phospholipase A2 brought maximum loss in enzyme activity suggested that this residue is involved in substrate binding (Chang et al. 1993).

The NBS-modified α-amylase from *Bacillus amyloliquifaciens* exhibited an increase in K_m of soluble starch. The latter result suggests that Trp residue is in a substrate-binding region (Kochhar and Dua 1985b). In glucokinase, however, there was no change in the K_m value of phosphoryl donors, but catalytic efficiency of enzyme decreased suggesting the involvement of Trp residue in catalysis (Hsieh et al. 1993).

4.5.5 Spectral Analysis

A decrease in absorbance at 280 nm was observed following inactivation of the enzyme upon reaction with NBS (Klein et al. 1985; Rawat and Rao 1996; Lee et al. 2000; Mardanyan et al. 2001; Kestwal and Bhide 2007; Gote et al. 2007). Fluorescence spectrum has been recorded in the modification of enzymes in the presence of NBS (Rawat and Rao 1996; Srinivasulu and Appu Rao 2000; Gote et al. 2007; Kestwal and Bhide 2007). In every case, the modified enzyme showed similar emission maximum as that of the native enzyme without any shift; however, quenching was observed (Rawat and Rao 1996; Srinivasulu and Appu Rao 2000; Gote et al. 2007; Kestwal and Bhide 2007). In the presence of a substrate, however, quenching was protected (Kestwal and Bhide 2007). CD spectral analysis of native and modified enzyme was also carried out. The CD spectra of native and modified enzymes were near identical, suggesting that the loss in activity was not because of any conformational change or extensive structural changes (Hsieh et al. 1993; Rawat

and Rao 1996; Lee et al. 2000). Difference spectra of modified α-galactosidase showed a decrease in absorbance along with isosbestic points around 260 and 305 nm that loss in activity is due to modification of Trp residue (Gote et al. 2007).

4.5.6 Protection Studies

Preincubation of the enzyme with the substrate or competitive inhibitor brought protection against NBS inactivation (Kochhar and Dua 1985b; Mathew and Balasubramaniam 1986). The modified glucokinase afforded protection in the presence of phosphoryl donors and xylose (Hsieh et al. 1993). The protection in the presence of substrates, competitive inhibitors indicate their binding to the enzyme active site.

4.6 Conclusion

The elucidation of functional groups (Cys, Ser, Tyr, and Trp) in the active site of an enzyme has been explained through the use of different group-specific reagents. When the enzyme is incubated with reagent, it may or may not result in loss of activity. The loss indicates the presence of that particular functional group. The pattern of loss in the activity was mostly single exponential (pseudo-first-order). In a few cases, biphasic kinetics was also observed. The modifications have also been monitored through UV-visible and fluorescence spectrum. The CD analysis of native and modified enzymes was very similar, suggesting that loss in activity is not associated with conformational change. Protection studies in the presence of substrates indicate the nature of their binding to active site and mechanism of catalysis.

References

Ambasht PK, Kayastha AM (1999) Plant Phosphoenolpyruvate phosphatase: a review. Physiol Mol Biol Plants 5:1–6

Ambasht PK, Malhotra OP, Kayastha AM (1996) Purification, characterization and steady state kinetic properties of cytosolic pyruvate kinase free of phosphoenol pyruvate phosphatase activity from germinating mung beans (Vigna radiata L.). Indian J Biochem Biophys 33:184–194

Ambasht PK, Malhotra OP, Kayastha AM (1997) Regulatory properties and active site groups of cytosolic mung bean pyruvate kinase. Indian J Biochem Biophys 34:365–372

Andrews AT, Reithel FJ (1970) The thiol groups of jack bean urease. Arch Biochem Biophys 141:538–546

Aspray TE, Riihimahi GM, Wolfe RG (1979) Malate dehydrogenase, inhibition of pig heart supernatant enzyme by iodoacetamide. J Biol Chem 254:1576–1579

Bednar RA (1990) Reactivity and pH dependence of thiol conjugation to *N*-ethylmaleimide—detection of a conformational change in chalcone isomerase. Biochemistry 29:3684–3690

Blanchard-Fillion B, Souza JM, Friel T, Jiang GCT, Vrana K, Sharav V, Barrion L, Schoneich C, Quizano C, Alvarez B, Radi R, Prezedborski S, Fernando GS, Horwitz J, Ischiropoulos H (2001) Nitration and inactivation of tyrosine hydroxylase by peroxynitrite. J Biol Chem 276:46017–46023

Britten CJ, Bird MI (1997) Chemical modification of an α 3-fructosyltransferase: definition of amino acid residues essential for enzyme activity. Biochim Biophys Acta 1334:57–64

Brocklehurst K (1979) Specific covalent modification of thiols: application in the study of enzymes and other biomolecules. Int J Biochem 10:259–274

Brown WE, Wold F (1973) Alkylisocyanates as active site specific reagents for serine proteases. Reaction properties. Biochemistry 12:828–834

Chang LS, Kuo KW, Chang CC (1993) Identification of functional involvement of tryptophan residues in phospholipase A2 from *Naja naja atra* (Taiwan cobra) snake venom. Biochim Biophys Acta 1202:216–220

Dairou J, Pluvinage B, Noiran J, Petit E, Vinh J, Haddad I, Mary J, Dupret J-M, Rodrigues-Lima M (2007) Nitration of a critical tyrosine residue in the allosteric inhibitor site of muscle glycogen phosphorylase impairs its catalytic activity. J Mol Biol 372:1009–1021

Duarte AR, Duarte DMR, Moreira KA, Cavalcanti MTH, de Lima-Filho JL, Porto AFL (2009) *Jacaratia corumbensis* O. Kuntz a new vegetable source of milk clotting enzymes. Braz Arch Biol Technol 52:1–9

Duff SMG, Lefebvre DD, Plaxton WC (1989) Purification and characterization of phosphoenol-pyruvate phosphatase from Brassica niger suspension cells. Plant Physiol 90:734–741

Ellman GL (1959) Tissue sulfhydryl groups. Arch Biochem Biophys 82:70–77

Eyzaguirre J (ed) (1987) Chemical modification of enzymes—an overview. The use of group-specific reagents. Ellis Horwood Ltd, Chichester, pp 9–22

Fatania HR, Al-Nassar KE, Thomas N (1993) Chemical modification of rat liver cytosolic $NADP^+$-linked isocitrate dehydrogenase by *N*-ethylmaleimide. FEBS Lett 322:245–248

Femfert U, Pfleiderer G (1969) The tyrosyl residues at the active site of aminopeptidase M, Modifications by tetranitromethane. FEBS Lett 4:262–264

Gorin G, Chin CC (1965) Urease IV: its reaction with *N*-ethylmaleimide and with silver ion. Biochim Biophys Acta 99:418–425

Gote MM, Khan MI, Khire JM (2007) Active site directed chemical modification of α-galactosidase from Bacillus stearothermophilus(NCIM 5146): involvement of lysine, tryptophan and carboxylate residues in catalytic site. Enzym Microb Technol 40:1312–1320

Granjeiro PA, Ferreira CV, Cavagis ADM, Granjeiro JM, Aoyama H (2003) Essential sulfhydryl groups in the active site of castor bean (Ricinus communis) seed acid phosphatase. Plant Sci 164:629–633

Hayashi S, Suzuki M, Nakamura S (1983) Chemical modification of Corynebacterium sarcosine oxidase: role of sulfhydryl and histidyl groups. J Biochem 94:551–558

Hsieh PC, Shenoy BC, Haase FC, Jentoft JE, Phillips NFB (1993) Involvement of tryptophan(s) at the active site of polyphosphate/ATP glucokinase from Mycobacterium tuberculosis. Biochemistry 32:6243–6249

Huang DY, Ichikawa Y (1995) Identification of essential lysyl and cysteinyl residues, and the amino acid sequence at substrate binding site of retinal oxidase. Biochim Biophys Acta 1243:431–436

Isaacs NS, Abed OH (1982) The mechanism of aromatic nitration by tetranitromethane. Tetrahydron Lett 23:2799–2802

Ischiripoulos H (2009) Protein tyrosine nitration—an update. Arch Biochem Biophys 484:117–121

Isla MI, Vattuone MA, Sampietro AR (1998) Essential groups at the active site of Trapaeolum invertase. Phytochemistry 47:1189–1193

Kaminska J, Wisniewska A, Koscielak J (2003) Chemical modification of alpha 1,6-fucosyltransferase define amino acid residues of catalytic importance. Biochimie 85:303–310

Kenyon GL, Bruice TW (1977) Novel sulfhydryl reagents. Methods Enzymol 47:407–430

Kestwal RM, Bhide SV (2007) Purification of β-galactosidase from Erythrina indica: involvement of tryptophan in active site. Biochim Biophys Acta 1770:1506–1512
Klein BP, Cohen BS, Grossman S, King D, Malovany H, Pinsky A (1985) Effect of modification of soybean lipoxygenase-1 with N-bromosuccinimide on linoleate oxidation, pigment bleaching and carbonyl production. Phytochemistry 24:1903–1906
Kochhar S, Dua RD (1985a) Chemical modification of liquefying α-amylase: role of tyrosine residues at its active center. Arch Biochem Biophys 240:757–767
Kochhar S, Dua RD (1985b) An active center tryptophan residue in liquefying α-amylase from *Bacillus amyloliquefaciens*. Biochem Biophys Res Commun 126:966–973
Komissarov AA, Debabov VG (1995) Modification with tetranitromethane of an essential tyrosine residue in uridine phosphorylase from Escherichia coli. Biochim Biophys Acta 1252:239–244
Kovina M, Viriyasov M, Baratova L, Kochetov G (1996) Localization of essential tyrosine residues of baker's yeast transketolase. FEBS Lett 392:293–294
Krishnan IS, Dua RD (1985) Clostridium pasteurianum glutamine synthetase mechanism. Evidence for active site residues. FEBS Lett 185:267–271
Kumar AK, Malhotra OP (1988) Active site groups; chemical inactivation of mung bean phosphoglycerate kinase. Plant Sci 56:113–116
Kumar S, Kayastha AM (2010) Soybean (Glycine max) urease: significance of sulfhydryl groups in urea catalysis. Plant Physiol Biochem 48:746–750
Kumari M, Sharma A, Jagannadham MV (2010) Decolorization of crude latex by activated charcoal, purification and physico-chemical characterization of religosin, a milk clotting serine protease from the latex of Ficus religosa. J Agric Food Chem 58:8027–8034
Kundu S, Sundd M, Jagannadham MV (2000) Purification and characterization of a stable cysteine protease Ervatamin B with two disulfide bridges, from the latex of Ervatamia coronaria. J Agric Food Chem 40:171–179
Lee WI, Fung H-L (2003) Mechanism-based partial inactivation of glutathione S-transferase by nitroglycerin. Tyrosine nitration versus sulfhydryl oxidation. Nitric Oxide 8:103–110
Lee YS, Kim HW, Lee KB, Park SS (2000) Involvement of arginine and tryptophan residues in catalytic activity of glutaryl 7-aminocephalosporanic acid acylase from Pseudomonas sp. strain GK 16. Biochim Biophys Acta 1523:123–127
Liu T-Y (1977) The role of sulphur in proteins. In: Neurath H, Hill RL (eds) The proteins. Academic Press, New York, pp 240–402
Lowe G (1976) The cysteine proteinases. Tetrahydron 32:291–302
Malhotra OP, Kayastha AM (1989) Chemical inactivation and active site groups of phosphoenolpyruvate phosphatase from germinating mung beans (Vigna radiata). Plant Sci 65:161–170
Malhotra OP, Kayastha AM (1990) Isolation and characterization of phosphoenolpyruvate phosphatase from germinating mung beans (Vigna radiata). Plant Physiol 93:194–200
Malhotra OP, Singh J, Srivastava PK (1991) Significance of sulphhydryl groups in the activity of isocitrate lyase of castor endosperm. Plant Sci 75:47–53
Malhotra OP, Srivastava DK, Kayastha AM (1993) Srinivasan. Inactivation of glyceraldehyde-3-phosphate dehydrogenase with SH-reagents and its relationship to the protein quaternary structure. Indian J Biochem Biophys 30:264–269
Maralihalli GB, Bhagwat AS (1992) Modification of maize phosphoenolpyruvate carboxylase by tetranitromethane. Phytochemistry 31:1529–1532
Mardanyan S, Sharoyan S, Antonyan A, Armenyan A, Cristalli G, Lupidi G (2001) Tryptophan environment in adenosine deaminase I. enzyme modification with *N*-bromosuccinimide in the presence of adenosine and EHNA analogues. Biochim Biophys Acta 1546:185–195
Maruthiah T, Esakkiraj P, Prabakaran G, Palavesam A, Immanuel G (2013) Purification and characterization of moderately halophilic alkaline serine protease from marine Bacillus subtilis AP-MSU 6. Biocatal Agric Biotechnol 2:116–119
Mathew CD, Balasubramaniam K (1986) Chemical modification of α-galactosidase from coconut. Phytochemistry 25:2439–2443

Matkawala F, Nighojkar S, Kumar A, Nighojkar A (2019) A novel thiol-dependent serine protease from *Neocosmospora* sp. N1. Heliyon 5:e02246

Murray AW, Froscio M, Rogers A (1974) Dissociation of rabbit muscle cyclic AMP-dependent protein kinase into catalytic and regulatory subunits by p-chloromercuribenzoate and methylmercuric chloride. FEBS Lett 48:238–240

Nagashima F, Tanase S, Morino Y (1986) Inactivation of cytosolic aspartate aminotransferase accompanying modification of Trp-48 by *N*-bromosuccinimide. FEBS Lett 197:129–133

Niu P, Yang K (2002) The 11-β hydroxysteroid dehydrogenase type 2 activity in human placental microsomes is inactivated by zinc and sulfhydryl group modifying reagent *N*-ethylmalimide. Biochim Biophys Acta 1594:364–371

Pande M, Dubey VK, Yadav SC, Jagannadham MV (2006) A novel serine protease cryptolepain from Cryptolepis buchanani: purification and biochemical characterization. J Agric Food Chem 54:10141–10150

Prakash O, Bhushan G (1998) A study of inhibition of urease from seeds of watermelon (Citrullus vulgaris). J Enzyme Inhib Med Chem 13:69–77

Rachadech W, Nimpiboon P, Naumthong W, Nakapong S, Krusong K, Pongsawasdi P (2015) Identification of essential tryptophan in amylomaltase from Corynebacterium glutamicum. Int J Biol Macromol 76:230–235

Ramchandran LK (1956) Protein iodine interactions. Chem Rev 56:199–218

Raskovic B, Bozovic O, Prodanovic R, Niketic V, Polovic N (2014) Identification, purification and characterization of a novel collegenolytic serine protease from fig (Ficus caricavar. Brown Turkey) latex. J Biosci Bioeng 118:622–627

Rawat UB, Rao MB (1996) Purification, kinetic characterization and involvement of tryptophan residue at the NADPH binding site of xylose reductase from Neurospora crassa. Biochim Biophys Acta 1293:222–230

Riordan JF, Vallee BL (1972a) Reactions with *N*-ethylmaleimide and p-mercuribenzoate. Methods Enzymol 25:449–456

Riordan JF, Vallee BL (1972b) Nitration with tetranitromethane. Methods Enzymol 25:515–521

Roknabadi SM, Bose SK, Taneja V (1999) A histidine thiol 150 kDa, tetrameric acid phosphatase from lentil, Lens esculenta, seeds with the characteristics of protein tyrosine phosphatases. Biochim Biophys Acta 1433:272–280

Singh AN, Shukla AK, Jagannadham MV, Dubey VK (2010) Purification of a novel cysteine protease, procerain B, from *Calotropis procera* with distinct characteristics compared to procerain. Process Biochem 45:399–406

Singh VK, Patel AK, Moir A-J, Jagannadham MV (2008) Indicain, a dimeric serine protease from *Morus indica* cv. K2. Phytochemistry 69:2110–2119

Skoubas A, Georgatos JG (1997) Identification of essential amino acids for the catalytic activity of barley β-glucosidase. Phytochemistry 46:997–1003

Sokolovsky M, Riordan JF (1970) On the use of tetranitromethane as a nitration regent for tyrosyl residues. FEBS Lett 9:241–243

Sokolovsky M, Riordan JF, Valle BL (1966) Tetranitromethane. A reagent for the nitration of tyrosyl residues in proteins. Biochemistry 5:3582–3589

Song KB, Swaisgood HG, Horton HR (1986) Requirement for a sulfhydryl group for sulfhydryl oxidase activity. J Dairy Sci 69:2589–2592

Souza JM, Radi M (1998) Glyceraldehyde-3-phosphate dehydrogenase inactivation by peroxnitrite. Arch Biochem Biophys 360:187–194

Spande TF, Witkop B (1967) Determination of tryptophan content of protein with *N*-bromosuccinimide. Methods Enzymol 11:498–506

Srinivasulu S, Appu Rao AG (2000) Role of tryptophan residues of lipooxygenase-1 in activity, structure and stability: chemical modification studies with *N*-bromosuccinimide. Food Chem 70:199–204

Srivastava PK, Kayastha AM (2000) Significance of the sulfhydryl groups in the activity of urease from pigeonpea (*Cajanus cajan* L.) seeds. Plant Sci 159:149–158

Stiborova M (1988) The role of cysteine and cysteine residues in phosphoenolpyruvate carboxylase from maize leaves. Biochem Physiol Pflanzen 183:7–14

Stratilova E, Dzurova M, Markovic O, Jornvall H (1996) An essential tyrosine residue of *Aspergillus* polygalacturonase. FEBS Lett 382:164–166

Stryer L (1995) Biochemistry, 4th edn. WH Freeman & Co, New York

Suzuki H (2015) In: Suzuki H (ed) Active site structure. How enzymes work from structure to function. Pan Stanford Publishing Pte Ltd, New York, pp 117–139

Suzuki H, Kawamura-Konishi Y (1991) Cysteine residues in the active site of Corynebacterium sarcosine oxidase. J Biochem 109:909–917

Tavakoli H, Ghourchian H, Moosavi-Moovaedi AA, Saboury AA (2006) Histidine and serine roles in catalytic activity of choline oxidase from *Alcaligenes* species studied by covalent modifications. Process Biochem 41:477–482

Todd MJ, Hausinger RP (1991) Identification of essential cysteine residue in *Klebsiella aerogenes* urease. J Biol Chem 266:24327–24331

Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91

Winther JR, Thorpe C (2014) Quantification of thiols and disulfides. Biochim Biophys Acta 1840:838–846

Worku Y, Luzio JP, Newby AC (1984) Identification of histidyl and cysteinyl residues essential for catalysis by 5′nucleotidase. FEBS Lett 167:235–240

Yadav SC, Pande M, Jagannadham MV (2006) Highly stable glycosylated serine protease from the medicinal plant *Euphorbia mili*. Phytochemistry 67:1414–1426

Yan JX, Kett WC, Herbert BR, Gooley AA, Packer NH, Williams KL (1998) Identification and quantitation of cysteine in proteins separated by gel electrophoresis. J Chromatogr A 813:187–200

Yang CC, Huang CS, Lee HJ (1985) Studies on the status of tyrosyl residues in phospholipase A2 from Naja naja atraand Naja nigricollis snake venoms. J Protein Chem 4:87–102

Yang SJ, Jiang SS, Tzeng CM, Kuo SY, Hung SH, Pan RL (1996) Involvement of tyrosine residue in the inhibition of plant vacuolar H^+-pyrophosphatase by tetranitromethane. Biochim Biophys Acta 1294:89–97

Zhang Q-S, Wang E-D, Wang Y-L (1998) The role of tryptophan residues in Escherichia coli arginyl-tRNA synthetase. Biochim Biophys Acta 1387:136–142

5 Use of Group-Specific Reagents in Active Site Functional Group Elucidation II: Asp, Glu, Arg, Lys, and His Residues

Pravin Kumar Ambasht

Abstract

The use of group-specific reagents for acidic and basic amino acid residue side chain have been discussed. The reagents include carbodiimide, Woodward's reagent K (Asp/Glu), phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione (Arg), trinitrobenzene sulfonic acid, pyridoxal phosphate (Lys), and diethylpyrocarbonate (His). The effect of incubation of different group-specific reagents on enzyme activity has been tested. In case of loss in activity, from inactivation kinetics data, values of $t_{1/2}$ and k are determined. The inactivation in the presence of substrates and inhibitors was also studied along with spectral analysis. This chapter will be helpful in understanding the modification of enzymes and the mechanism of the reaction.

Keywords

Carbodiimide · Phenylglyoxal · Trinitrobenzenesulfonic acid · Pyridoxal phosphate · Photooxidation · Diethylpyrocarbonate · Inactivation kinetics · Protection

5.1 Introduction

In the previous chapter, I discussed the use of group-specific reagents for the identification of active site functional groups like Cys, Ser, Tyr, and Trp. In the present chapter, the focus is on the identification of acidic and basic amino acid

P. K. Ambasht (✉)
Department of Biochemistry, School of Life Sciences, North-Eastern Hill University, Shillong, Meghalaya, India
e-mail: pambasht@nehu.ac.in

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_5

residues like Asp, Glu, Arg, Lys, and His. Among these, Asp is the most acidic and Arg, the most basic amino acids. The role of pK_a is, therefore, very important. In addition to the use of group-specific reagents, photooxidation has also been discussed for the identification of His residues. Protection to the loss of activity in the presence of reagents proved to help understand the nature of their binding.

5.2 Aspartic Acid and Glutamic Acid

Some enzymes have a carboxylic acid side chain (Asp and Glu) as part of the active site. Asp is a stronger acid in comparison to Glu. Asp is part of the active site of chymotrypsin, while Asp and Glu are together involved in the catalysis of lysozyme (Stryer 1995).

5.2.1 Characteristics of Reagents and Methods

A method for modification of the carboxyl group was developed using water-soluble carbodiimide (Hoare and Koshland 1967). The chemistry of carbodiimide has been extensively discussed earlier (Khorana 1953). Carbodiimide brings about activation of the carboxyl group. Commonly used carbodiimide derivatives included BDC (1-benzyl-3-dimethyl aminopropyl-carbodiimide), EDC (1-ethyl-3-dimethyl aminopropyl carbodiimide), and CMC (1-cyclohexyl-3-(2-morpholino ethyl) carbodiimide). The activated carbodiimide is allowed to react with a nucleophile like a glycine methyl ester. Variations in the structure of carbodiimide and variations in the properties of nucleophiles can alter the type of modification of the carboxyl group. Further, the method works well in the presence of urea and guanidine hydrochloride (Hoare and Koshland 1967). Glycine ethyl ester has also been used (Singh and Kayastha 2014). The reaction between enzyme and EDC has been carried out at pH 6.0 (Dua and Gupta 1985a; Gote et al. 2007; Singh and Kayastha 2014). EDC, in addition to carboxylic acid, also modifies SH and Tyr residues (Singh and Kayastha 2014).

Another reagent used for the identification of carboxylic acid in the active site is Woodward's reagent K (*N*-ethyl-5-phenyl-isoxazolium-3′-sulfonate). The reaction of the reagent with enzyme was carried out in 50 mM sodium phosphate buffer, pH 6.0 (Gote et al. 2007). The reaction of Woodward's reagent K with invertase was, however, carried out in 200 mM MES buffer, pH 4.7 (Isla et al. 1998). The reagent has been reported to be very unstable in an aqueous medium (Dunn et al. 1974). Different reports on the use of group-specific reagents for the identification of carboxylic acid have been summarized in Table 5.1. Limited literature is available on this aspect. The structures of different group-specific reagents have been shown in Fig. 5.1.

Table 5.1 Use of group-specific reagents for the identification of carboxylic acid in the active site

S. no.	Reagents	Enzyme	Reference
1.	Carbodiimide	Carboxypeptidase A	Dua and Gupta (1985a)
2.		α-Galactosidase	Mathew and Balasubramaniam (1986), Gote et al. (2007)
3.		α-Amylase	Singh and Kayastha (2014)
4.	Woodward's reagent K	Nuclease	Dunn et al. (1974)
5.		Invertase	Isla et al. (1998)
6.		α-Galactosidase	Gote et al. (2007)

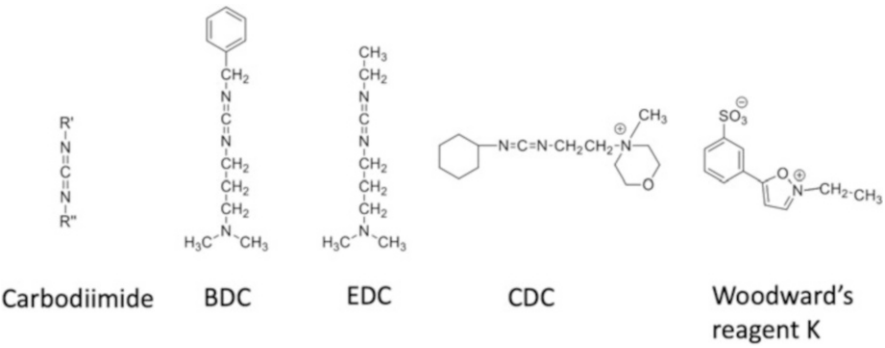


Fig. 5.1 Structures of different group-specific reagents for the identification of carboxylic acid in the active site

5.2.2 Inactivation Kinetics

The presence of a carboxyl group in the active site of coconut α-galactosidase and wheat α-amylase was established using EDC (Mathew and Balasubramaniam 1986; Singh and Kayastha 2014). When the enzymes were incubated in the presence of EDC, it resulted in the loss of activity, following single exponential decay or pseudo-first-order kinetics. In both the above enzymes, protection of loss in activity was observed in the presence of substrates (Mathew and Balasubramaniam 1986; Singh and Kayastha 2014). Goat pancreas carboxypeptidase, when incubated with CMC, resulted in the loss of enzyme activity. The loss in activity was pH and CMC concentration-dependent. The experiment was conducted between pH 8.0 and 6.0. The optimal loss occurred at pH 6.0 in the presence of 50 mM CMC. Single exponential decay was observed in all cases. The plot of k versus [CMC] exhibited a nonlinear plot suggesting a two-step reaction. CMC-modified carboxypeptidase A was desalted, and the time-dependent loss in peptidase and esterase activities were

monitored. The loss of activity runs parallel, indicating that the carboxyl group is involved in catalysis (Dua and Gupta 1985a).

In wheat α-amylase, pK_a value 4.87 indicates the presence of carboxylic acid as part of the active site (Singh and Kayastha 2014). From the plot of log inactivation rate versus [Carbodiimide], the slope was determined and found to be 1.8, suggesting that two carboxyl groups are at or near active site (Mathew and Balasubramaniam 1986).

In the presence of Woodward's reagent K, invertase from *Tropaeolum* and α-galactosidase from *Bacillus stearothermophilus* exhibited a loss in activity. The results suggested the presence of carboxylic acid in the active site of the above enzymes (Isla et al. 1998; Gote et al. 2007). The latter enzyme also exhibited a loss in activity in the presence of EDC (Gote et al. 2007). In the case of α-galactosidase, the inactivation was concentration-dependent and did not follow first-order kinetics. Kinetic analysis revealed that K_m value remains unchanged (Gote et al. 2007). The nuclease activity was lost in the presence of Woodward's reagent K and was pH and reagent concentration-dependent (Dunn et al. 1974).

Substrates and competitive inhibitors have shown to bring about protection to the loss of activity. Sucrose brought protection to the loss of activity, but fructose did not in *Tropaeolum* invertase (Isla et al. 1998). β-Phenylpropionic acid and glycyl-L-tyrosine exhibited protection to both peptidase and esterase activities in CMC-modified carboxypeptidase A (Dua and Gupta 1985a). The CD spectra of modified and native enzyme (α-galactosidase) were identical, suggesting that modification did not bring any gross conformational change (Gote et al. 2007).

5.3 Arginine

Arginine is the most basic amino acid with the presence of the guanidino group. It functions in the catalysis of different enzymes being part of the active site. Arg residues have a general role as anionic binding sites (Riordan 1979). Different reagents used for the modification and identification of Arg residue are phenylglyoxal, 2, 3-butanedione, and 1, 2-cyclohexanedione.

5.3.1 Methods

Phenylglyoxal reacts with Arg under mild conditions between pH 7.0 and 8.0 at 25 °C to yield a derivative containing two phenylglyoxal per guanidine group of arginine (Takahashi 1968). Other reagents like 2, 3-butanedione, and 1, 2-cyclohexanedione react with Arg at pH 8.0 and 8.5, respectively (Kang et al. 1997; Chang et al. 2004). 1,2-Cyclohexanedione condenses with guanidine group of Arg to form cyclohexanedione-Arg (CHD-Arg) under alkaline aqueous medium (Toi et al. 1967). The use of phenylglyoxal for the identification of Trp residue has

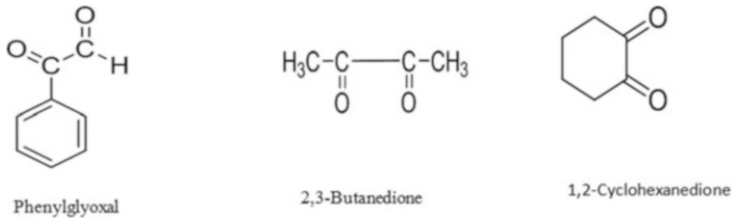


Fig. 5.2 Structures of group-specific reagents for the identification of Arg residue

also been reported (Kojima et al. 1991). Phenylglyoxal and its derivatives like hydroxyphenylglyoxal, chlorophenylglyoxal, dimethylaminophenylglyoxal, and *O*-methylphenylglyoxal were tested for their effects on *N*-acetylarginine and buffers as a probe for Arg investigation (Baburaj and Durrani 1991). *p*-Hydroxyphenylglyoxal reacts with Arg residue under mild conditions between pH 7.0 and 9.0 at 25 °C. The extent of modification can be determined at 340 nm using the molar absorption coefficient of $1.83 \times 10^4\ M^{-1}\ cm^{-1}$ (Yamasaki et al. 1980). The structures of commonly used group-specific reagents for the identification of Arg residues have been shown in Fig. 5.2.

5.3.2 Reports on the Identification of Arg in an Active Site Using Different Group-Specific Reagents

In most reports, phenylglyoxal and 2, 3-butanedione have been used. Red beet plasma membrane Ca^{2+} ATPase and mung bean vacuolar H^+-PPase were found to be more sensitive to phenylglyoxal than butanedione (Basu and Briskin 1996; Hsiao et al. 2007). Phenylglyoxal is more hydrophobic than 2, 3-butanedione; it is suggested that Arg is embedded in a more hydrophobic environment (Hsiao et al. 2007). Because of the modification of Arg residue in plasma membrane Ca^{2+} ATPase, both ATP-dependent Ca^{2+} transport and ITP hydrolysis rate were inhibited (Basu and Briskin 1996). Rubisco has been extensively studied with respect to the identification of Arg residue in the presence of phenylglyoxal and 2, 3-butanedione (Schloss et al. 1978; Chollet 1981; Mizohata et al. 2003). In the presence of 1, 2-cyclohexanedione, reversible modification of Arg residues was reported with native egg white lysozyme and bovine pancreatic RNase (Patthy and Smith 1975). Reports of the use of group-specific reagents for Arg have been summarized in Table 5.2.

5.3.3 Inactivation Kinetics

Different enzymes exhibited inactivation when allowed to react with Arg-specific reagents with loss of enzyme activity (Dua and Gupta 1985b; Baijal and Sane 1988; Kang et al. 1997; Kaminska et al. 2003; Chen et al. 2005). The inactivation in

Table 5.2 Reports on the investigation of Arg residue in an active site using group-specific reagents

S. no.	Reagents	Enzyme	Reference
1.	Phenylglyoxal	*Ustilago* RNase U1	Hashimoto and Takahashi (1977)
2.		Ribulose bisphosphate carboxylase	Schloss et al. (1978), Chollet (1981)
3.		Porcine phospholipase A2	Vensel and Kantrowitz (1980)
4.		Porcine pancreatic phospholipase A2	Fleer et al. (1981)
5.		Goat carboxypeptidase A	Dua and Gupta (1985b)
6.		*Escherichia coli* alkaline phosphatase	Chaidaroglou et al. (1988)
7.		*Amaranthus* nitrate reductase complex	Baijal and Sane (1988)
8.		Mung bean seedling pyrophosphatase	Kuo and Pan (1990)
9.		*Zea mays* malic enzyme	Rao et al. (1991)
10.		Pig kidney diamine oxidase	Shah et al. (1996)
11.		Red beet plasma membrane Ca^{2+} ATPase	Basu and Briskin (1996)
12.		*E. coli* aldolase	Qamar et al. (1996)
13.		Taiwan cobra phospholipase A2	Chang et al. (1998)
14.		*Pseudomonas* 7-aminocephalo-sporanic acid acylase	Lee et al. (2000)
15.		Spinach Rubisco	Mizohata et al. (2003)
16.		Platelet α1,6-fucosyltransferase	Kaminska et al. (2003)
17.		*Notechis scutatus* phospholipase A2	Chang et al. (2004)
18.		*Pinctada fucata* alkaline phosphatase	Chen et al. (2005)
19.		Mung bean vacuolar H^+pyrophosphatase	Hsiao et al. (2007)
20	2,3-Butanedione	Porcine heart malate dehydrogenase	Bleile et al. (1975)
21.		Porcine phospholipase A2	Vensel and Kantrowitz (1980)
22.		Ribulose bisphosphate carboxylase	Chollet (1981)
23.		*Amaranthus* nitrate reductase complex	Baijal and Sane (1988)
24.		Mung bean seedling pyrophosphatase	Kuo and Pan (1990)
25.		*Zea mays* malic enzyme	Rao et al. (1991)
26.		*E. coli* aldolase	Qamar et al. (1996)

(continued)

Table 5.2 (continued)

S. no.	Reagents	Enzyme	Reference
27.		Red beet plasma membrane Ca^{2+} ATPase	Basu and Briskin (1996)
28.		Pig kidney diamine oxidase	Shah et al. (1996)
29.		*Pisum sativum* malate dehydrogenase	Kang et al. (1997)
30.		Human phenol sulfotransferase	Chen and Chen (2003)
31.		Mung bean vacuolar H^+pyrophosphatase	Hsiao et al. (2007)
32.	1,2-Cyclohexane-dione	Bovine pancreatic RNase	Patthy and Smith (1975), Richardson et al. (1990)
33.		Egg white lysozyme	Patthy and Smith (1975)
34.		Porcine phospholipase A2	Vensel and Kantrowitz (1980)
35.		Porcine pancreatic phospholipase A2	Fleer et al. (1981)
36.		Taiwan cobra phospholipase A2	Chang et al. (1998)
37.		*Notechis scutatus* phospholipase A2	Chang et al. (2004)
38.	Phosphoric acid mono (2,3-Dioxo-butyl) ester	Rabbit muscle aldolase	Chabot et al. (2008)

$NADP^+$-dependent malate dehydrogenase from *Pisum sativum* leaves was time and concentration-dependent (Kang et al. 1997). Pig kidney diamine oxidase in the presence of 2, 3-butanedione exhibited photochemical inactivation (50% loss in the activity was observed in 30 min) in the presence of light but did not show a loss in activity under dark condition. In the presence of phenylglyoxal, a 50% loss in activity was observed only (Shah et al. 1996). The loss of activity in red beet plasma membrane Ca^{2+} ATPase in the presence of 2, 3-butanedione and phenylglyoxal exhibited nonlinear log % residual activity versus time plot (Basu and Briskin 1996).

Single first-order (pseudo-first-order) kinetics was observed in some cases (Schloss et al. 1978; Vensel and Kantrowitz 1980; Dua and Gupta 1985b; Rao et al. 1991; Chang et al. 2004). Second-order kinetics of inactivation has also been observed (Chen et al. 2005). Nitrate reductase complex (NADH-NR, $FADH_2$-NR, and NADH dehydrogenase) from *Amaranthus dubious* leaves when subjected to treatment with phenylglyoxal or 2, 3-butanedione brought a loss of activity in a biphasic manner. The initial loss was rapid, while later, it was slow (Baijal and Sane 1988). Phenylglyoxal modified a single Arg residue of carboxypeptidase from the goat pancreas with an increase in absorbance at 250 nm (Dua and Gupta 1985b). Porcine phospholipase A2 exhibited maximum loss of activity in the presence of phenylglyoxal. The rate of inactivation of the enzyme was reagent concentration-dependent. Further, pH also influenced the rate of reaction between enzyme and

phenylglyoxal; with an increase in pH from 6.5 to 9.5, the second-order *k* increased almost ten times (Vensel and Kantrowitz 1980). Cytoplasmic malate dehydrogenase from porcine heart muscle exhibited inactivation (pseudo-first-order kinetics) in the presence of 2, 3-butanedione with modification of four Arg residues. In the case of the formation of the ternary complex (E.NADH.Hydroxymalonate), all four Arg were inaccessible for modification (Bleile et al. 1975). Pyrophosphatase from etiolated mung bean in the presence of phenylglyoxal and 2, 3-butanedione lost the activity following pseudo-first-order kinetics. The lower $t_{1/2}$ and K_i for phenylglyoxal suggest that it is more effective than 2, 3-butanedione (Kuo and Pan 1990). Human phenol sulfotransferase in the presence of 2, 3-butanedione exhibited time- and concentration-dependent inactivation. The inactivation followed pseudo-first-order kinetics (Chen and Chen 2003).

5.3.4 Protection Studies

Protection studies have also been carried out (Rao et al. 1991; Basu and Briskin 1996; Lee et al. 2000; Mizohata et al. 2003). The inactivation of glutaryl 7-aminocephalosporanic acid acylase (heterotetramer) from *Pseudomonas* could not be protected by glutaric acid, 7-aminocephalosporanic acid (competitive inhibitor), and substrate glutaryl 7-aminocephalosporanic acid (Lee et al. 2000). Some other enzymes inactivated in the presence of Arg-specific reagents, exhibited protection in the presence of substrate or competitive inhibitor (Schloss et al. 1978; Baijal and Sane 1988; Rao et al. 1991; Basu and Briskin 1996; Kang et al. 1997; Mizohata et al. 2003; Chen et al. 2005). NADH exhibited protection in the slow phase inactivation of nitrate reductase complex (NADH-NR-NADH dehydrogenase). The results suggest that NR contains Arg residue in the active site responsible for the binding of NADH (Baijal and Sane 1988). In malic enzyme from maize (*Zea mays*), protection in the presence of NADP alone or combination with Mg^{2+} ion and malate was observed, suggesting that Arg residues are located at or near the active site. The modification showed that the enzyme had a similar affinity for NADPH; however, the affinity for malate was reduced. This suggests that Arg residue is involved in the binding of malate (Rao et al. 1991). ATP exhibited protection to plasma membrane Ca^{2+} ATPase suggesting the binding of ATP to the active site, with some conformational change making the Arg residue less susceptible to the attack of 2, 3-butanedione and phenylglyoxal (Basu and Briskin 1996). The modification of the Arg residue of goat pancreas carboxypeptidase A was protected in the presence of β-phenylpropionic acid (Dua and Gupta 1985b). GDP and GDP-fucose protected against inactivation of α1,6-fucosyltransferase from human platelets in the presence of phenylglyoxal (Kaminska et al. 2003). In the case of a modification of Arg residue in the presence of phenylglyoxal of spinach, Rubisco brought inactivation, but the product 3-PGA protected inactivation (Mizohata et al. 2003). The cytoplasmic malate dehydrogenase exhibits protection to two Arg residues out of four in the presence of NADH. In the presence of AMP, protection from inactivation was observed but not in the presence of nicotinamide. The result suggests that the

AMP moiety of NADH is responsible for protection (Bleile et al. 1975). In the case of etiolated mung bean pyrophosphatase, substrates provided partial protection against these reagents. The overall result is suggestive of the presence of at least one Arg residue in the active site (Kuo and Pan 1990). In most cases, it is being suggested that Arg residue is critical and is located in the active site and is involved in the binding of a substrate (Dua and Gupta 1985b; Kang et al. 1997; Kaminska et al. 2003; Chang et al. 2004; Chen et al. 2005).

5.3.5 Site-Directed Mutagenesis

The effects of phenylglyoxal on wild-type and mutant (Arg 166 → Ala) alkaline phosphatase from *Escherichia coli* was investigated. It was found that wild type exhibited inactivation, while in the mutant enzyme no loss in activity was observed. The result suggests the presence of Arg residue in the active sites (Chaidaroglou et al. 1988).

5.3.6 Modification and Conformation

The inactivation of Diamine oxidase in the presence of phenylglyoxal was without any conformational change; however, when 10 Arg residues were modified, conformational change was observed suggesting their structural role as well (Shah et al. 1996). Rubisco was shown to lose activity during activity assay despite the presence of large excess of substrate known as "Fallover." It is being explained by slow conformational change-dependent loss in the activity. However, modification of Arg residues led to a reduction in the fallover phenomenon (Mizohata et al. 2003).

5.4 Lysine

Lysine is another basic amino acid. It has primary amine as the side chain and is involved in the catalytic functions in different enzymes. Several reagents have been used for modification and identification of Lys residue in the active site. Pyridoxal phosphate (PLP) and trinitrobenzenesulfonic acid (TNBS) are commonly used.

5.4.1 Reagent Characteristics and Methods

PLP reacts with ε-NH_2 of Lys residue to form Schiff's base, and its reduction in the presence of $NaBH_4$ results in a loss in activity (Benesch et al. 1972; Huang and Ichikawa 1995). The reaction of PLP with ε-NH_2 of Lys residue of the enzyme was studied at pH 7.4. The molar extinction coefficient of the reaction of PLP with Lys ε-NH_2 was 9.72×10^3 M^{-1} cm^{-1} at 325 nm (Huang and Ichikawa 1995). TNBS does not exhibit absorbance between 320 and 480 nm. The trinitrophenyl-enzyme

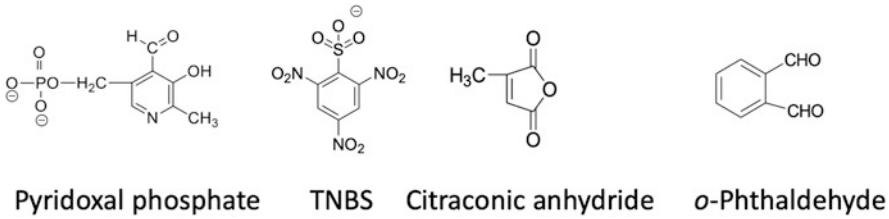


Fig. 5.3 Structures of group-specific reagents for identification of Lys residue

Fig. 5.4 Structure of trinitrophenyl-lysine

shows a broad peak between 345–348 nm and a shoulder at 420 nm. The extinction coefficient at 346 nm of trinitrophenyl-Lys is 1.45×10^4 M^{-1} cm^{-1}. The experiments are carried out in the dark because trinitrophenyl-Lys is photosensitive (Hollenberg et al. 1971).

Citraconic anhydride (2-Methylmaleic anhydride) brings about citraconylation reaction and has been used for the identification of Lys residue in the active site (Gibbons and Perham 1970; Gote et al. 2007). The citraconylation reaction was performed at pH 8.2 (Gibbons and Perham 1970). The structure of different reagents employed for the identification of Lys residues in the active site is shown in Fig. 5.3**.** The modification of Lys in the presence of TNBS results in the formation of trinitrophenyl-lysine. The structure of the latter is shown in Fig. 5.4. There are several reports on the identification of Lys residue using group-specific reagents and have been summarized in Table 5.3.

5.4.2 Inactivation Kinetics

Different enzymes exhibited inactivation when incubated with PLP/$NaBH_4$ (Wimmer and Harrison 1975; Bleile et al. 1976; Pandey and Iyengar 2002; Kaminska et al. 2003). Porcine heart mitochondrial malate dehydrogenase and

Table 5.3 Identification of Lys residue in an active site using different group-specific reagents

S. no.	Reagent	Enzyme	Reference
1.	Pyridoxal phosphate	Porcine heart mitochondrial malate dehydrogenase	Wimmer and Harrison (1975)
2.		Porcine heart cytoplasmic malate dehydrogenase	Bleile et al. (1976)
3.		Bovine pancreatic RNase A	Richardson et al. (1990)
4.		Chicken liver PEP-carboxykinase	Guidinger and Nowak (1991)
5.		Snake venom Notexin	Chang (1996)
6.		Rat mevalonate kinase	Potter et al. (1997)
7.		*Enterobacter aerogens* glycerol dehydrogenase	Pandey and Iyengar (2002)
8.		Platelet α-1,6-Fucosyl transferase	Kaminska et al. (2003)
9.	TNBS	Bovine liver glutamate dehydrogenase	Coffee et al. (1971)
10.		Rabbit muscle pyruvate kinase	Hollenberg et al. (1971)
11.		Calf liver glutathione reductase	Carlberg and Mannervik (1979)
12.		Snake venom phospholipase A2	Yang and Chang (1989)
13.		Pig kidney diamine oxidase	Shah et al. (1996)
14.		*Pinctada fucata* alkaline phosphatase	Chen et al. (2005)
15.	Citraconic anhydride	Rabbit muscle aldolase	Gibbons and Perham (1970)
16.		*Ustilago sphaerogena* Ribonuclease U1	Hashimoto and Takahashi (1977)
17.		*Bacillus stearothermophilus* α-galactosidase	Gote et al. (2007)
18.	4-chloro-3,5-dinitrobenzoate	Snake venom phospholipase A2	Chang et al. (1994)
19.	2,4-Dinitrophenyl-propionate	*Clostridium acetobutylicum* acetoacetate decarboxylase	Schmidt and Westheimer (1971)
20.	*o*-Phthalaldehyde	*Enterobacter aerogens* glycerol dehydrogenase	Pandey and Iyengar (2002)
21.	2,4-pentanedione	Chicken liver PEP-carboxykinase	Guidinger and Nowak (1991)
22.	Iodoacetamide	*Corynebacterium* sp. U-96 sarcosine oxidase	Mukouyama et al. (2004)

cytoplasmic malate dehydrogenase exhibited reversible inactivation in the presence of pyridoxal-5′ phosphate due to the formation of Schiff's base with loss in activity (Wimmer and Harrison 1975; Bleile et al. 1976). When Schiff's base was subjected to a reduction in the presence of $NaBH_4$, irreversible inactivation was observed

(Wimmer and Harrison 1975). A biphasic kinetic pattern was observed with the initial phase representing the reversible formation of Schiff's base and the second phase representing the irreversible formation of X-azolidine-like structure. The biphasic plot had two pseudo-first-order rates (Bleile et al. 1976).

Phospho*enol*pyruvate carboxykinase from the chicken liver was modified in the presence of 2, 4-pentanedione and pyridoxal-5′-phosphate. The inactivation of enzyme in the presence of 2, 4-pentanedione followed pseudo-first-order kinetics and was linearly concentration-dependent. In the presence of pyridoxal-5′phosphate inactivation follows bimolecular kinetics. The p*K* value determined was 8.1. The overall result suggests the presence of one reactive Lys at the active site (Guidinger and Nowak 1991). Notexin from snake venom in the presence of pyridoxal-5′phosphate modified 2 Lys residues out of 11 residues, namely Lys-82 and Lys-115. The modified notexin exhibited an increase in activity (Chang 1996).

Glycerol dehydrogenase from *Enterobacter aerogens* was inactivated in the presence of pyridoxal 5′-phosphate and *o*-phthalaldehyde due to modification. The inhibition was time and concentration-dependent. The *o*-phthalaldehyde modified the proximal Lys residue and a Cys residue with enzyme inactivation (pseudo-first-order kinetics). In the latter modification, there is the formation of a thioisoindole derivative as evident from emission maximum at 415 nm. (Pandey and Iyengar 2002).

There are reports on the inactivation of different enzymes due to their modification in the presence of TNBS (Coffee et al. 1971; Hollenberg et al. 1971). Bovine liver glutamate dehydrogenase exhibited modification of two Lys residues upon treatment with TNBS, namely Lys-425 and Lys-428. The latter reacts rapidly in comparison to Lys-425. The latter is modified only when Lys-428 has been modified (Coffee et al. 1971). Rabbit muscle pyruvate kinase inactivation followed pseudo-first-order kinetics. The modified enzyme upon hydrolysis gave ε-*N*-trinitrophenyllysine (Hollenberg et al. 1971).

Phospholipase A2 from snake venom of Taiwan cobra exhibited a loss in activity and resulted in the formation of two triphenylated derivative TNP-1 and TNP-2. TNP-1 contained Lys-6, while TNP-2 contained Lys-6 as well as Lys-65. The reactivity of Lys-6 and Lys-65 was enhanced towards TNBS in the presence of Ca^{2+}. These results suggest that two Lys residues were not involved in Ca^{2+} binding (Yang and Chang 1989). Diamine oxidase from pig kidney upon treatment with TNBS exhibited two kinds of Lys residues. Around 21 Lys residues did not show any loss in activity and changes in conformation, while the remaining 19 Lys residues showed changes in hydrodynamic properties and loss in activity (Shah et al. 1996). Alkaline phosphatase from *Pinctada fucata* was inactivated due to the modification of Lys residue and the loss in activity followed pseudo-first-order kinetics (Chen et al. 2005). The inactivation reaction with TNBS of α-galactosidase from *Bacillus stearothermophilus* was concentration-dependent and did not show pseudo-first-order kinetics (Gote et al. 2007).

Phospholipase A2 from snake venom exhibited a loss in the activity upon treatment with 4-chloro-3, 5-dinitrobenzoate. The reaction resulted in the carboxydinitrophenylation of Lys-6 (Chang et al. 1994). Acetoacetate decarboxylase from *Clostridium acetobutylicum* in the presence of 2, 4-dinitrophenyl propionate was modified. A Lys amino group was considered to be important for the catalytic mechanism. The loss of activity followed pseudo-first-order kinetics. The p*K* value determined was 5.9, almost four p*K* units less than usual. The result suggests that it represents actual p*K* of an amino group at the active site (Schmidt Jr. and Westheimer 1971).

Iodoacetamide is usually used for the identification of a thiol group. There is however a report of modification of Lys residue in the presence of iodoacetamide. Sarcosine oxidase from *Corynebacterium* sp. U-96 exhibited inactivation in the presence of iodoacetamide and followed pseudo-first-order kinetics (Mukouyama et al. 2004).

5.4.3 Protection Studies

The Schiff base formed when porcine heart mitochondrial malate dehydrogenase and cytoplasmic malate dehydrogenase were subjected to PLP showed protection in the presence of NADH (Wimmer and Harrison 1975; Bleile et al. 1976). The protection result suggests that the interaction of pyridoxal phosphate was near the active site (Bleile et al. 1976). Chicken liver PEP-carboxykinase modified in the presence of PLP exhibited complete protection of inactivation in the presence of IDP, ITP, and PEP (Guidinger and Nowak 1991). NAD^+ and NADH exhibited protection against inactivation of PLP-modified glycerol dehydrogenase. The result suggested the presence of active Lys residue near the coenzyme-binding site. Glycerol partially protected the inactivation of *o*-phthalaldehyde-modified glycerol dehydrogenase but NAD^+ was ineffective. The Lys involved in *o*-phthalaldehyde inactivation was different from pyridoxal-5′-phosphate inactivated lysine (Pandey and Iyengar 2002). GDP-fucose and GDP offered protection against inactivation in the presence of pyridoxal phosphate/$NaBH_4$of α-1, 6-fucosyl transferase from human platelets. The result suggests that Lys residue is close to the substrate-binding site (Kaminska et al. 2003).

There are reports on the protection of enzymes inactivated in the presence of TNBS. ADP exhibited strong protection against inactivation to rabbit muscle pyruvate kinase but PEP did not. Further, ATP also protected the activity against inactivation, in a similar way to ADP. The results suggest that Lys residue is important for the binding of ADP and catalysis (Hollenberg et al. 1971). Substrate binding had little effect on alkaline phosphatase inactivation (Chen et al. 2005). Sarcosine oxidase from *Corynebacterium* sp. U-96 modified in the presence of iodoacetamide exhibited protection in the presence of acetate buffer of pH 6.4, 7.2 and 8.0 (Mukouyama et al. 2004).

5.4.4 Spectral Analysis

CD spectral results indicated little change in the secondary structure of pyridoxal-phosphate-modified chicken liver PEP-carboxykinase (Guidinger and Nowak 1991). CD spectra of the PLP-modified notexin did not show any change in the secondary structure and with respect to its affinity for Ca^{2+} suggesting that Lys residue is not involved in calcium-binding (Chang 1996). The fluorescence spectral data of the PLP-modified and sodium borohydride-reduced glycerol dehydrogenase indicated modification of the ε-NH_2 group of a specific Lys residue (Pandey and Iyengar 2002). CD spectroscopic results suggest that inactivation did not bring any conformational change to TNBS-modified α-galactosidase, and hence loss in the activity is not structural change related. The result suggests that Lys residue takes part in the catalysis as well as substrate binding (Gote et al. 2007). The modification of phospholipase A2 in the presence of 4-chloro-3, 5-dinitrobenzoate did not bring any significant change in the secondary structure and Ca^{2+} binding. The result suggests that Lys-6 plays an important role in the mechanism of catalysis of phospholipase A2 (Chang et al. 1994).

5.5 Histidine

Histidine is a basic amino acid containing side-chain imidazole group. This is part of the active site of several enzymes. Its role in active site of chymotrypsin is well explained along with the use of tosyl L-phenylalanine-chloromethylketone (TPCK) (Stryer 1995). The role of His in enzyme-active site has been reviewed (Miles 1977; Schneider 1978). The identification of His residue has been approached through the determination of pK_a value, photooxidation, and use of diethylpyrocarbonate (DEPC).

5.5.1 Methods

Rate versus pH plot helps in the determination of pK_a value and gives a tentative indication of the functional group (Ambasht et al. 1997). Photooxidation in the presence of dyes methylene blue and rose Bengal has been discussed for identification of His. Methylene blue is cationic, while rose Bengal is anionic. Methylene blue shows low specificity as some other residues like Trp, Tyr, Ser, and Thr are also modified (Bellin and Yankus 1968). Rose Bengal is more specific for the imidazole group at neutral pH (Westhead 1972). In the experiment, the enzyme in the presence of a light source (100 W bulb) is kept at 30 cm and illuminated. The time-dependent loss in activity is monitored. The control experiment includes enzyme and dye in the dark; enzyme alone in the dark, and in light (Malhotra and Kayastha 1989).

Diethylpyrocarbonate (DEPC) reacts with His residue to form carbethoxy-histidyl derivative (Lundblad 2014). The reaction is pH-sensitive, i.e., it is highly specific at near-neutral pH. In the alkaline pH, the reaction is reversed (Eyzaguirre

Table 5.4 Reports of determination of pK_a value in relation to His residue in enzyme-active site

S. no.	Enzyme	pK_a	Reference
1.	Malate dehydrogenase	7.0	Aspray et al. (1979)
2.	PEP-phosphatase	7.2	Malhotra and Kayastha (1990)
3.	Pyruvate kinase	6.6	Ambasht et al. (1997)
4.	H^+ ATPase	6.7	Chu et al. (2001)
5.	Urease	6.2	Srivastava and Kayastha (2001)
6.	Alkaline phosphatase	6.6	Ding et al. (2002)
7.	Pyrophosphatase	6.4	Hsiao et al. (2002)
8.	α-Amylase	6.8	Singh and Kayastha (2014)

1987). The loss in activity is followed by an increase in absorbance at 240 nm. Hydroxylamine removes the carbethoxy-group from *N*-carbethoxyhistidyl residue (Miles 1977). The treatment with hydroxylamine leads to the opening of the imidazole ring with de-ethoxyformylation by the Bamberg reaction (Srivastava and Kayastha 2001). DEPC is also shown to modify the imidazole ring of purine to its carbethoxy form, and also, the ring can be cleaved by the Bamberger reaction (Loosemore and Pratt 1976). The reaction of DEPC with His residue takes place when both nitrogen atoms of His are free. The reports on the determination of pK_a for His residue have been summarized in Table 5.4. It is evident that pK_a values were between 6.2 and 7.2 representing perturbed values for His.

5.5.2 Photooxidation

Single exponential loss in activity (pseudo-first-order kinetics) was observed in almost every case suggesting the presence of the imidazole group and its importance for catalysis (Dua and Kochhar 1985; Malhotra and Singh 1992; Ambasht et al. 1997; Isla et al. 1998). The photooxidized enzyme exhibited an increase in absorbance at 250 nm that was directly proportional to the extent of inactivation (Dua and Kochhar 1985). Substrates have been shown to bring protection against photoinactivation (Dua and Kochhar 1985; Isla et al. 1998; Srivastava and Kayastha 2001). The results of photooxidation have been summarized in Table 5.5.

5.5.3 DEPC

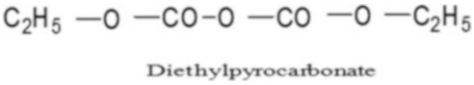
An attempt was made with N of His residue, coordinated through Ru (II) or Pt (II) in a metalloprotein or metalloenzyme, but the reaction did not take place (Jackman et al. 1988). However, there is a report of reaction between DEPC with Co(III) complexes [Co(III)(en)Cl His]Cl and [Co(III)(dien)His]Cl_2 (Li and Rosenberg 1993). The structure of DEPC has been shown in Fig. 5.5.

Jack bean urease displays aggregation in the presence of Cu^{2+}; however, upon modification with DEPC, the enzyme's affinity for Cu^{2+} decreases and reduces

Table 5.5 Results of photooxidation in the exploration of His residue in the active site

S. no.	Dye	Enzyme	Reference
1.	Methylene blue	Mung bean PEP-phosphatase	Malhotra and Kayastha (1989)
2.		Mung bean phosphoglycerate kinase	Kumar and Malhotra (1988)
3.		Mung bean pyruvate kinase	Ambasht et al. (1997)
4.		*Tropaeolum* invertase	Isla et al. (1998)
5.		Pigeon pea urease	Srivastava and Kayastha (2001)
6.	Rose Bengal	*Bacillus amyloliquefaciens* α-amylase	Dua and Kochhar (1985)
7.		Mung bean PEP-phosphatase	Malhotra and Kayastha (1989), Kayastha and Malhotra (1993)
8.		Castor isocitrate lyase	Malhotra et al. (1987), Malhotra and Singh (1992)
9.		Mung bean pyruvate kinase	Ambasht et al. (1997)
10.		Pigeon pea urease	Srivastava and Kayastha (2001)

Fig. 5.5 Structure of DEPC



aggregation. Jack bean urease contains 24 His residues; however, DEPC brings about modification of 18–20 residues. Four His residues are coordinated with Ni in the active site and are not accessible to DEPC as only 35% loss in activity was observed. The CD spectra of native and DEPC-modified enzyme were very similar, suggesting that secondary structure was retained (Follmer and Carlini 2005). The results of the use of DEPC in the modification of His residue in the active site have been summarized in Table 5.6.

5.5.3.1 Kinetics

Inactivation of enzyme activity has been observed in the presence of DEPC (Khan and McFadden 1982; Hsiao et al. 2002; Colleluori et al. 2005; Tavakoli et al. 2006). In most reports, single exponential loss in activity (first-order kinetics) has been observed (Dua and Kochhar 1985; Malhotra and Singh 1992; Ambasht et al. 1997; Roknabadi et al. 1999). Isocitrate lyase from flax indicated modification of two His residue per subunit of the tetrameric enzyme (Khan and McFadden 1982). A similar result was obtained with castor isocitrate lyase as well (Malhotra and Singh 1992). α-3-Fucosyl transferase showed inactivation with IC_{50} (92 μM) and complete inhibition at 500 μM (Britten and Bird 1997).The reaction of $NADP^+$-malate dehydrogenase from *Pisum sativum* and α-amylase from wheat in the presence of DEPC was time and concentration-dependent (Kang et al. 1997; Singh and Kayastha 2014). Choline oxidase from *Alcaligenes* exhibited a loss in activity in the presence

Table 5.6 Results of use of DEPC in the exploration of His residue in the active site

S. no.	Enzyme	Reference
1.	*Escherichia intermedia* tyrosine phenol lyase	Kumagai et al. (1975)
2.	Flax isocitrate lyase	Khan and McFadden (1982)
3.	*Rhodospirillum rubrum* F1-ATPase	Khananshivili and Gromet-Elhanan (1983)
4.	*Corynebacterium* sarcosine oxidase	Hayashi et al. (1983)
5.	Cytoplasmic 5′nucleotidase	Worku et al. (1984)
6.	Maize δ-aminolevulinic acid dehydratase	Maralihalli et al. (1985)
7.	Watermelon isocitrate lyase	Jameel et al. (1985)
8.	*Bacillus amyloliquefaciens* α-amylase	Dua and Kochhar (1985)
9.	Castor isocitrate lyase	Malhotra and Singh (1992)
10.	Bovine inositol monophosphatase	Rees-Milton et al. (1993)
11.	Porcine kidney aminopeptidase P	Lim and Turner (1996)
12.	α3-Fucosyl transferase	Britten and Bird (1997)
13.	*Pisum sativum* malate dehydrogenase	Kang et al. (1997)
14.	Mung bean pyruvate kinase	Ambasht et al. (1997)
15.	Barley β-glucosidase	Skoubas and Georgatos (1997)
16.	*Lens esculenta* acid phosphatase	Roknabadi et al. (1999)
17.	*Phaseolus vulgaris* arginase	Carvazal et al. (1997)
18.	*Tropaeolum* invertase	Isla et al. (1998)
19.	Pigeonpea urease	Srivastava and Kayastha (2001)
20.	Vacuolar H^+-ATPase	Chu et al. (2001)
21.	*Megalobatrachus japonicus* alkaline phosphatase	Ding et al. (2002)
22.	α1,6-Fucosyl transferase	Kaminska et al. (2003)
23.	Rat liver arginase	Colleluori et al. (2005)
24.	*Alcaligenes* choline oxidase	Tavakoli et al. (2006)
25.	Wheat α-amylase	Singh and Kayastha (2014)
26.	*Rigidoporus lignosus* laccase	Vianello et al. (2014)

of DEPC, and the latter was a competitive inhibitor as indicated in the Lineweaver-Burk plot (Tavakoli et al. 2006). The presence of DEPC brought changes to V_{max} only and not K_m in vacuolar H^+ pyrophosphatase from etiolated mung beans (Hsiao et al. 2002).

In one case, DEPC was found to modify His residue in isozymes of phospho-*enol*pyruvate carboxylase from maize in such a way that there is the dissociation of the tetrameric structure. PEPCI is more sensitive in comparison to PEPCII towards reaction with DEPC (Stiborova 1989). Laccase from *Rigidoporus lignosus* was modified with DEPC, which revealed the presence of a His residue in the active site, based on kinetic behavior and mass spectrometry results (Vianello et al. 2014).

5.5.3.2 Effect of Hydroxylamine

Restoration of activity has been reported in the presence of hydroxylamine (Khan and McFadden 1982; Stiborova 1989; Lim and Turner 1996; Isla et al. 1998).

Hydroxylamine, in certain cases, provided partial restoration or not any significant recovery (Srivastava and Kayastha 2001; Chu et al. 2001). The restoration of activity in the presence of hydroxylamine was different with respect to isozymes of 5′nucleotidase. Hydroxylamine completely reversed the inhibition of the cytosolic enzyme, while partial reversal of inhibition was observed in membrane 5′nucleotidase (Worku et al. 1984).

5.5.3.3 Protection Studies

Substrates have shown to bring protection against DEPC inactivation (Dua and Kochhar 1985; Maralihalli et al. 1985; Srivastava and Kayastha 2001; Hsiao et al. 2002). Acetate exhibited complete protection to the DEPC-modified sarcosine oxidase (Hayashi et al. 1983). In the case of arginase from rat liver, L-ornithine, and borate together brought complete protection while L-ornithine, alone brought partial protection (Colleluori et al. 2005). In some other cases, the substrate did not show protection (Britten and Bird 1997). NADPH (10 mM) exhibited more than 80% restoration of activity of chloroplast $NADP^+$-malate dehydrogenase from pea leaves (Kang et al. 1997).

Glycerol-3-phosphate, AMP, and Pi protected the alkaline phosphatase activity, suggesting that modified His in the presence of DEPC is located in the active site (Ding et al. 2002). L-Ala shows protection against inactivation (Kumagai et al. 1975). GDP-fucose did not bring any protection against inactivation of DEPC-modified α1,6-fucosyltransferase (Kaminska et al. 2003). Succinate and glyoxylate exhibited protection against inactivation due to DEPC in flax and watermelon isocitrate lyase (Khan and McFadden 1982; Jameel et al. 1985). Succinate and glyoxylate, when present together, resulted in better protection (Jameel et al. 1985). No protection in the presence of glyoxylate, however, was observed in castor isocitrate lyase (Malhotra and Singh 1992).

5.5.3.4 Spectral Analysis

The reaction between DEPC and various enzymes results in inactivation. The loss in activity was with a corresponding increase in absorbance at 240 nm (Malhotra and Singh 1992; Ambasht et al. 1997; Chu et al. 2001). In some other reports, an increase in absorbance was recorded at 242 nm (Kumagai et al. 1975; Ding et al. 2002; Colleluori et al. 2005). The increase in absorbance was without any change in absorbance at 278 nm, suggesting the presence of His residue at the active site and ruling out the presence of Tyr residue (Malhotra and Singh 1992; Ambasht et al. 1997; Roknabadi et al. 1999). The increase in absorbance at 240 nm was also lost in the presence of hydroxylamine (Maralihalli et al. 1985; Hsiao et al. 2002; Singh and Kayastha 2014). The fluorescence spectrum of DEPC bound to choline oxidase displayed quenching in the intensity. The Far-UV CD spectrum reveals that modification of His residue is responsible for the change in secondary structure (Tavakoli et al. 2006). The extent of modification of sarcosine oxidase was determined by difference absorption spectrum, the peaks were at 246 and 292 nm and isosbestic points at 237 and 283 nm (Hayashi et al. 1983).

5.5.4 Iodoacetamide

Iodoacetamide is primarily used for modification of the thiol group. However, there is a report on the modification of His residue in the presence of iodoacetamide through alkylation. Malate dehydrogenase from pig heart was modified, giving rise to 1,3-dicarboxamidomethyl histidine, 3-carboxamidomethyl histidine, and 1-carboxamidomethyl histidine. 1,3-Dicarboxamidomethyl histidine is responsible for the inactivation of the enzyme. Protection against inactivation was observed in the presence of NADH binding. The overall result suggests the presence of essential His residue at or near the active site. The inactivation followed pseudo-first-order kinetics (Aspray et al. 1979).

5.6 Conclusion

The challenge of understanding of enzyme-active site functional groups has been sorted out by the use of different group-specific reagents. Different kinds of spectra like UV visible, fluorescence and CD, inactivation kinetics, and protection in the presence of substrates have also contributed to the understanding. With the advancement, a technique like site-directed mutagenesis has brought strength in the exploration of active site functional groups.

References

Ambasht PK, Malhotra OP, Kayastha AM (1997) Regulatory properties and active site groups of cytosolic mung bean pyruvate kinase. Indian J Biochem Biophys 34:365–372

Aspray TE, Riihimahi GM, Wolfe RG (1979) Malate dehydrogenase, inhibition of pig heart supernatant enzyme by iodoacetamide. J Biol Chem 254:1576–1579

Baburaj K, Durrani S (1991) Exploring arylglyoxals as the arginine reactive probes. A mechanistic investigation using the buffer and substituent effects. Bioorg Chem 19:229–244

Baijal M, Sane PV (1988) Arginine residue(s) at the active site(s) of the nitrate reductase complex from *Amranthus*. Phytochemistry 27:1969–1972

Basu S, Briskin DP (1996) Chemical modification of essential arginine residues associated with the red beet (Beta vulgaris L.) plasma membrane Ca^{2+}-ATPase. Phytochemistry 41:685–689

Bellin JS, Yankus CA (1968) Influence of dye binding on the sensitized photooxidation of amino acids. Arch Biochem Biophys 123:18–28

Benesch RE, Benesch R, Renthal RD, Maeda N (1972) Affinity labeling of the polyphosphate binding site of hemoglobin. Biochemistry 11:3576–3582

Bleile DM, Foster M, Brady JW, Harrison JH (1975) Identification of essential arginyl residues in cytoplasmic malate dehydrogenase with butanedione. J Biol Chem 250:6222–6227

Bleile DM, Jameson JL, Harrison JH (1976) Inactivation of porcine heart cytoplasmic malate dehydrogenase by pyridoxal 5′phosphate. J Biol Chem 251:6304–6307

Britten CJ, Bird MI (1997) Chemical modification of an α-3-fructosyltransferase: definition of amino acid residues essential for enzyme activity. Biochim Biophys Acta 1334:57–64

Carlberg I, Mannervik B (1979) Inhibition of glutathione reductase by interaction of 2,4,6-trinitrobenzene sulfonate with the active-site dithiol. FEBS Lett 98:263–266

Carvazal N, Enriquez S, Salas M, Uribe E (1997) A critical histidine residue in arginase from *Phaseolus vulgaris*. Phytochemistry 46:1327–1329

Chabot N, Vinatier V, Gefflaut T, Baudoin C, Rodriguez F, Blonski C, Hoffmann P (2008) Irreversible inhibition of aldolase by a phosphorylated α-dicarbonyl compound. J Enzyme Inhib Med Chem 23:21–27

Chaidaroglou A, Brezinski DA, Middleton SA, Kantrowitz ER (1988) Function of arginine-166 in the active site of Escherichia coli alkaline phosphatase. Biochemistry 27:8338–8343

Chang LS (1996) Chemical modification of notexin from Notechis scutatus scutatus (Australian tiger snake) venom with pyridoxal-5′phosphate. J Protein Chem 15:473–480

Chang LS, Hung JJ, Lin SR, Chang CC (1994) Chemical modification of Lys-6 in Taiwan cobra phospholipase A2 with 4-chloro-3,5-dinitrobenzoate. Biochem Mol Biol Int 33:1207–1213

Chang LS, Lin SR, Chang CC (1998) Identification of Arg-30 as the essential residue for the enzymatic activity of Taiwan cobra phospholipase A2. J Biochem 124:764–768

Chang LS, Wu PF, Liou JC, Chiang-Lin WH, Yang CC (2004) Chemical modification of arginine of Notechis scutatus scutatus notexin. Toxicon 44:491–497

Chen G, Chen X (2003) Arginine residues in the active site of human phenol sulfotransferase (SULT1A1). J Biol Chem 278:36358–36364

Chen H-T, Xi L-P, Yu Z-Y, Xu G-R, Zhang RQ (2005) Chemical modification studies on alkaline phosphatase from pearl oyster (*Pinctada fucata*): a substrate reaction course analysis and involvement of essential arginine and lysine residues at the active site. Int J Biochem Cell Biol 37:1446–1457

Chollet R (1981) Inactivation of crystalline tobacco ribulosebisphosphate carboxylase by modification of arginine residues with 2,3-butanedione and phenylglyoxal. Biochim Biophys Acta 658:177–190

Chu CL, Hsiao YY, Chen CH, Van RC, Lin WJ, Pan RL (2001) Inhibition of plant vacuolar H^+ ATPase by diethylpyrocarbonate. Biochim Biophys Acta 1505:12–22

Coffee CJ, Bradshaw RA, Goldin BR, Frieden C (1971) Identification of sites of modification of bovine liver glutamate dehydrogenase reacted with trinitrobenzenesulfonate. Biochemistry 10:3516–3526

Colleluori DM, Reczkowski RS, Emig FA, Cama E, Cox JD, Scolnick LR, Campher K, Jude K, Han S, Viola RE, Christianson DW, Ash DE (2005) Probing the role of hyper-reactive histidine residue of arginase. Arch Biochem Biophys 444:15–26

Ding S, Li Y, Zhu L (2002) Identification of histidine residues at the active site of *Megalobatrachus japonicus* alkaline phosphatase by chemical modification. Biochim Biophys Acta 31:100–108

Dua RD, Gupta K (1985a) Mechanistic studies on carboxypeptidase A from goat pancreas-Part II: evidence for carboxyl group. J Biosci 9:91–97

Dua RD, Gupta K (1985b) Mechanistic studies on carboxypeptidase A from goat pancreas: role of arginine residue at the active site. Arch Biochem Biophys 236:479–486

Dua RD, Kochhar S (1985) Active site studies on Bacillus amyloliquefaciens α-amylase (I). Mol Cell Biochem 66:13–20

Dunn BM, Anfinsen CB, Shrager BI (1974) Kinetics of Woodward's reagent K hydrolysis and reaction with Staphylococcal nuclease. J Biol Chem 249:3717–3723

Eyzaguirre J (1987) In: Eyzaguirre J (ed) Chemical modification of enzymes—an overview. The use of group-specific reagents. Ellis Horwood Ltd., Chichester, pp 9–22

Fleer EAM, Puijk WC, Slotboom AJ, Dehaas GH (1981) Modification of arginine residues in porcine pancreatic phospholipase A2. Eur J Biochem 116:277–284

Follmer C, Carlini CR (2005) Effect of chemical modifications of histidines on the copper-induced oligomerization of jackbean urease (EC 3.5.1.5). Arch Biochem Biophys 435:15–20

Gibbons I, Perham RN (1970) The reaction of aldolase with 2-methylmaleic anhydride. Biochem J 116:843–849

Gote MM, Khan MI, Khire JM (2007) Active site directed chemical modification of α-galactosidase from Bacillus stearothermophilus (NCIM 5146): involvement of lysine, tryptophan and carboxylate residues in catalytic site. Enzym Microb Technol 40:1312–1320

Guidinger PF, Nowak T (1991) An active-site lysine in avian liver phosphoenolpyruvate carboxykinase. Biochemistry 30:8851–8861

Hashimoto J, Takahashi K (1977) Chemical modification of ribonuclease U1. Biochem J 81:1175–1180

Hayashi S, Suzuki M, Nakamura S (1983) Chemical modification of Corynebacterium sarcosine oxidase: role of sulfhydryl and histidyl groups. J Biochem 94:551–558

Hoare DG, Koshland DE Jr (1967) A method for quantitative modification and estimation of carboxylic acid group in proteins. J Biol Chem 242:2447–2453

Hollenberg PF, Flashner M, Coon MJ (1971) Role of lysyl ε-amino groups in adenosine diphosphate binding and catalytic activity of pyruvate kinase. J Biol Chem 246:946–953

Hsiao YY, Pan YJ, Hsu SH, Huang YT, Liu TH, Lee CH, Lee CH, Liu PF, Chang WC, Wang YK, Chien LF, Pan RL (2007) Functional roles of arginine residues in mung bean vacuolar H^+ pyrophosphatase. Biochim Biophys Acta 1767:965–973

Hsiao YY, Van RC, Hung HH, Pan RL (2002) Diethylpyrocarbonate inhibition H^+-pyrophosphatase possibly involves a histidine residue. J Protein Chem 21:51–58

Huang DY, Ichikawa Y (1995) Identification of essential lysyl and cysteinyl residues, and the amino acid sequence at substrate binding site of retinal oxidase. Biochim Biophys Acta 1243:431–436

Isla MI, Vattuone MA, Sampietro AR (1998) Essential groups at the active site of Trapaeolum invertase. Phytochemistry 47:1189–1193

Jackman MP, Lim MC, Oswath P, de Silva Harshani DGA, Skykes AG (1988) Diagnostic test for ruthenium and platinum modification of histidine residues on metalloproteins using diethylpyrocarbonate (DEPC). Inorg Chim Acta 153:205–208

Jameel S, El-Gul T, McFadden BA (1985) Modification of active site of isocitrate lyase from watermelon. Arch Biochem Biophys 236:72–81

Kaminska J, Wisniewska A, Koscielak J (2003) Chemical modification of alpha 1,6-fucosyltransferase define amino acid residues of catalytic importance. Biochimie 85:303–310

Kang H-C, Hwang Y-S, Chung I-S, Cho K-J, Hahn T-R (1997) Essential arginyl and histidyl residues at the active site of NADP-malate dehydrogenase from Pisum sativum L. giant leaves. J Plant Physiol 150:497–503

Kayastha AM, Malhotra OP (1993) Studies on active site of a phosphoenolpyruvate phosphatase from germinating seeds of Vigna radiata L. Wilczek. Plant Physiol Biochem 20:24–27

Khan FR, McFadden BA (1982) Isocitrate lyase from flax: terminal residues, composition, active site and catalysis. Plant Physiol 70:943–948

Khananshivili D, Gromet-Elhanan Z (1983) Modification of histidine residues by diethylpyrocarbonate leads to inactivation of the Rhodospirillum ruberum RrF1-ATPase. FEBS Lett 159:271–274

Khorana HG (1953) The chemistry of carbodiimides. Chem Rev 53:145–166

Kojima E, Kai M, Ohkura Y (1991) Phenylglyoxal as a fluorogenic reagent selective for tryptophan. Anal Chim Acta 248:213–217

Kumagai H, Utagawa T, Yamada H (1975) Studies on tyrosine phenol lyase. Modification of essentials histidyl residues by diethylpyrocarbonate. J Biol Chem 250:1661–1667

Kumar AK, Malhotra OP (1988) Active site groups; chemical inactivation of mung bean phosphoglycerate kinase. Plant Sci 56:113–116

Kuo SY, Pan RL (1990) An essential arginine residue in pyrophosphatase of hypocotyls from etiolated mung beans. Plant Physiol 93:1128–1133

Lee YS, Kim HW, Lee KB, Park SS (2000) Involvement of arginine and tryptophan residues in catalytic activity of glutaryl 7-aminocephalosporanic acid acylase from Pseudomonas sp. strain GK 16. Biochim Biophys Acta 1523:123–127

Li C, Rosenberg RC (1993) Carbethoxylation of coordinated histidine by diethylpyrocarbonate. J Inorg Biochem 51:727–735

Lim J, Turner AJ (1996) Chemical modification of porcine kidney aminopeptidase P indicates the involvement of two critical histidine residues. FEBS Lett 381:188–190

Loosemore MJ, Pratt RF (1976) The irreversible cleavage of histidine residues by diethylpyrocarbonate (Ethoxyformyl anhydride). FEBS Lett 72:155–158
Lundblad RL (2014) Chemical reagents for protein modification, 4th edn. CRC Press, Taylor and Francis Group, Boca Raton, FL
Malhotra OP, Dwivedi UN, Singh J, Srivastava PK (1987) Chemical reaction mechanism and active-site groups of isocitrate lyase. Indian J Biochem Biophys 24:57–62
Malhotra OP, Kayastha AM (1989) Chemical inactivation and active site groups of phosphoenolpyruvate phosphatase from germinating mung beans (Vigna radiata). Plant Sci 65:161–170
Malhotra OP, Kayastha AM (1990) Isolation and characterization of phosphoenolpyruvate phosphatase from germinating mung beans (Vigna radiata). Plant Physiol 93:194–200
Malhotra OP, Singh J (1992) Role of histidine residues and mechanism of action of isocitrate lyase of castor seedling endosperm. Plant Sci 81:155–162
Maralihalli GB, Rao SR, Bhagwati AS (1985) Histidine residues at the active site of maize δ-aminolevulinic acid dehydratase. Phytochemistry 24:2533–2536
Mathew CD, Balasubramaniam K (1986) Chemical modification of α-galactosidase from coconut. Phytochemistry 25:2439–2443
Miles EW (1977) Modification of histidyl residues in proteins by diethylpyrocarbonate. Methods Enzymol 47:431–442
Mizohata E, Anwaruzzaman M, Okuno H, Tomizawa K-I, Shigeoka S, Kai Y, Yokota A (2003) Chemical modification of arginine alleviates the decline in activity during catalysis of spinach Rubisco. Biochem Biophys Res Commun 301:591–597
Mukouyama EB, Oguchi M, Kodera Y, Maeda T, Suzuki H (2004) Low pK_a lysine residues at the active site of sarcosine oxidase from Corynebacterium sp. U-96. Biochem Biophys Res Commun 320:846–851
Pandey A, Iyengar L (2002) Chemical modification of specific active site amino acid residues of Enterobacter aerogenes glycerol dehydrogenase. J Enzyme Inhib Med Chem 17:1749–1753
Patthy L, Smith EL (1975) Identification of a functional arginine residue in ribonuclease a and lysozyme. J Biol Chem 250:565–569
Potter D, Wojnar JM, Narasimham C, Miziorko HM (1997) Identification and functional characterization of an active site lysine in mevalonate kinase. J Biol Chem 272:5741–5746
Qamar S, Marsh K, Berry A (1996) Identification of arginine 331as an important active site residue in the Class II fructose-1,6-bisphosphate aldolase of Escherichia coli. Protein Sci 5:154–161
Rao SR, Kamath BG, Bhagwat AS (1991) Chemical modification of the functional arginine residue (s) of malic enzyme from Zea mays. Phytochemistry 30:431–435
Rees-Milton KJ, Greasley PJ, Ragan CI, Gore MG (1993) Bovine inositol monophosphatase. The identification of a histidine residue reactive to diethylpyrocarbonate. FEBS Lett 321:37–40
Richardson RM, Pares X, Cuchillo CM (1990) Chemical modification by pyridoxal 5′phosphate and cyclohexane 1,2-dione indicates the Lys-7 and Arg-10 are involved in the P2 phosphate-binding subsite of bovine pancreatic ribonuclease A. Biochem J 267:593–599
Riordan JF (1979) Arginyl residues and anion binding sites in proteins. Mol Cell Biochem 26:71–110
Roknabadi SM, Bose SK, Taneja V (1999) A histidine thiol 150 kDa, tetrameric acid phosphatase from lentil, Lens esculenta, seeds with the characteristics of protein tyrosine phosphatases. Biochim Biophys Acta 1433:272–280
Schloss JV, Norton IL, Stringer CD, Hartman FC (1978) Inactivation of ribulosebisphosphate carboxylase by modification of arginine residues with phenylglyoxal. Biochemistry 17:5626–5631
Schmidt DE Jr, Westheimer FH (1971) pK of lysine amino group at the active site of acetoacetate decarboxylase. Biochemistry 10:1249–1253
Schneider F (1978) Histidine in enzyme active centers. Angew Chem 17:583–592
Shah MS, Tayyab S, Ali R (1996) Probing structure-activity relationship in diamine oxidase-reactivities of lysine and arginine residues. Int J Biol Macromol 18:77–81

Singh K, Kayastha AM (2014) α-Amylase from wheat (Triticum aestivum) seeds: its purification, biochemical attributes and active site studies. Food Chem 162:1–9
Skoubas A, Georgatos JG (1997) Identification of essential amino acids for the catalytic activity of barley β-glucosidase. Phytochemistry 46:997–1003
Srivastava PK, Kayastha AM (2001) Studies on the histidine residues in pigeonpea (Cajanus cajan L.) urease. J Mol Catal B Enzym 16:81–89
Stiborova M (1989) The different responses of two isozymes of phosphoenolpyruvate carboxylase from maize (Zea mays L.) leaves to the modification by diethylpyrocarbonate. Biochem Physiol Pflanzen 184:227–234
Stryer L (1995) Biochemistry, 4th edn. WH Freeman & Co, New York
Takahashi K (1968) The reaction of phenylglyoxal with arginine residues in proteins. J Biol Chem 243:6171–6179
Tavakoli H, Ghourchian H, Moosavi-Moovaedi AA, Saboury AA (2006) Histidine and serine roles in catalytic activity of choline oxidase from Alcaligenes species studied by covalent modifications. Process Biochem 41:477–482
Toi K, Bynum E, Norris E, Itano HA (1967) Studies on the chemical modification of arginine. I. The reaction of 1,2-cyclohexanedione with arginine and arginyl residues of proteins. J Biol Chem 242:1036–1043
Vensel LA, Kantrowitz ER (1980) An essential arginine residue in porcine phospholipase A2. J Biol Chem 255:7306–7310
Vianello F, Miotto G, Cambria MT, Lima GPP, Vanzani P, Paolo MLD (2014) Kinetic role of a histidine residue in the T1 copper site of the laccase from Rigidosporous lignosus. J Mol Catal B Enzym 99:34–42
Westhead EW (1972) Dye sensitized photooxidation. Methods Enzymol 25:401–409
Wimmer MJ, Harrison JH (1975) Identification of an essential lysine in porcine heart mitochondrial malate dehydrogenase. J Biol Chem 250:8768–8773
Worku Y, Luzio JP, Newby AC (1984) Identification of histidyl and cysteinyl residues essential for catalysis by 5′nucleotidase. FEBS Lett 167:235–240
Yamasaki RB, Vega A, Feeney RE (1980) Modification of available arginine residues in proteins by p-Hydroxyphenylglyoxal. Anal Biochem 109:32–40
Yang CC, Chang LS (1989) Studies on the status of lysyl residues in phospholipase A2 from Naja naja atra (Taiwan cobra) snake venom. Biochem J 262:855–860

Protein-Protein Interactions Modeling: From Dry to Wet Lab

6

Ekta Khare and Dev Bukhsh Singh

Abstract

The basic study of protein provides insight into their role in health and diseases. Protein–protein interaction (PPI) networks are intrinsic to virtually every cellular process and of vital importance in systems biology. Besides the knowledge of protein three-dimensional structures, their dynamics and interaction profiles are important to completely understand the biochemical mechanisms at the molecular level and utilize it for the benefit of mankind. Protein–protein interactions can alter the kinetic properties of proteins, and this can be reflected in the form of crucial alteration in specificity to substrates, catalysis, or altered allosteric properties of the complex. Understanding of protein complex formation allows an analysis of molecular functions, and hence needed to redesign and produce tailor-made proteins suitable for desired work. In the last decades, a vast amount of PPI data was generated by high-throughput experimental methods. However, voluminous unexplored proteins, creating the computational analysis of the PPI network a mandatory tool. Moreover, each method carries its advantages and disadvantages to reproduce a true PPI model that precisely set up under physiological conditions. To properly understand the significance of PPI in the cell, one needs to thoroughly verify through the physical, molecular biological, and genetic approaches that have been used to detect protein–protein interactions. This chapter describes the various wet and dry lab approaches to study the molecular details of interactions and to consider the reliability of each method.

E. Khare (✉)
Department of Microbiology, Institute of Biosciences and Biotechnology, Chhatrapati Shahu Ji Maharaj University, Kanpur, Uttar Pradesh, India

D. B. Singh
Department of Biotechnology, Institute of Biosciences and Biotechnology, Chhatrapati Shahu Ji Maharaj University, Kanpur, Uttar Pradesh, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_6

Keywords

Protein–protein interaction · In silico techniques · In vitro techniques · In vivo techniques

6.1 Introduction

Protein–protein interactions (PPIs) are the mandatory part of every single life on earth from unicellular to multicellular, prokaryotes to eukaryotes, and plants to animals. It is impossible to understand cellular functions without a thorough knowledge of PPIs. Moreover, facts associated with PPIs are also important for designing of new generation drugs. A protein molecule in itself is made up of a specific sequence of amino acids that fold into secondary, tertiary, and quaternary structures under the influence of posttranslational modifications, characteristic pH, and redox environment of the cell. Furthermore, understanding the interactions between proteins is thus even a more challenging process due to the lack of complete information on proteins and gene expression events for the genome of organisms.

Over 80% of proteins have been estimated to operate as complexes (Berggård et al. 2007). Well-known examples of multi-subunit proteins are RNA polymerases, pyruvate dehydrogenase, etc. Many proteins–protein interactions are part of larger cellular networks of cellular pathways. A proteome illustration of interactions between proteins will provide unparalleled information to figure out the development, cellular communications, and biochemistry of the cell (Zhang 2009). Protein interactions can be categorized by functional and structural properties (Zhang 2009). The important characteristics for understanding the function of PPIs are life span, stability, and affinity. PPIs can be characterized by the life span as permanent or transient and based on stability into obligate or non-obligate. Transient protein complexes can be further differentiated into strong or weak in terms of affinity (Nooren and Thornton 2003; Braun and Gingras 2012). Several proteins are known to exist as components of permanent obligate complexes like multi-subunit. Non-obligate complexes involved in enzyme–inhibitor, enzyme–substrate, hormone–receptor, and signaling–effector types of interactions are of short duration enzymes (Mintseris and Weng 2005).

Protein complexes of long duration formed due to strong interactions. Classical biochemical methods such as size exclusion chromatography or native gel electrophoresis can be used for the assessment of such strong obligate interactions (Podobnik et al. 2016). Traditional in vitro and in vivo approaches are not appropriate for detection and analysis of transient interactions but require extremely responsive and high-resolution experimental techniques (Sali et al. 2003). Experimental methods like co-immunoprecipitation or affinity chromatography are often laborious and expensive for the determination of obligate and non-obligate or transient interactions. The outputs of these experiments might be affected by system errors (Geva and Sharan 2010). It becomes a challenge to validate all the data of protein–protein interactions in laboratory generated by high-throughput researches of the

century. Interactions with the protein of known function can determine the role of unidentified protein. Computational approaches are of great use to understand the functions of the unexplored protein. Keeping these notions into consideration, efficient use of integrative computational approaches is indispensable to thoroughly investigate the protein–protein interactions. The goal of this chapter is to present various methods from wet and dry lab to rationalize tools for the investigation of protein interactions of various nature and strength.

6.2 Preferred Mode of Protein Associations

The process of association formation began with the random search of protein at a rate, as stated by Brownian motion. This is trailed by accurate docking of the protein interfaces. Still, a collision between proteins does not authenticate the formation of the complex as their accurate relative orientation is necessary for electrostatic attraction to initiate the process of protein–protein interaction and complex formation (Selzer and Schreiber 2001). The interface formed between the pair of proteins plays an important role in the interaction. The amino acid residues of interacting proteins contact in this interfacial region (Tuncbag et al. 2008). The interfaces between permanently interacting proteins are in general larger in size (ranging from 1500 to 10,000 $Å^2$) over transient interfaces (<1500 $Å^2$) (De et al. 2005; Park et al. 2009). The amino acid composition of transient interfaces is not very much different from the rest of the protein surface; however, it shows a somewhat higher number of neutral polar groups (Ansari and Helms 2005; Perkins et al. 2010). This is because each component of this non-obligate complex has to subsist separately in the cell. Strongly interacting proteins have interfaces rich in hydrophobic sites so that upon interaction they can form more widespread conformational changes (Janin et al. 2008).

Hydrophobic interactions are the leading force in protein coupling driven by a gain in free energy of binding (Tripathi 2013). However, this energy is not evenly distributed across the interfaces. Recent studies generalized that on interfaces, a small number of residues mainly contribute to the binding free energy of protein–protein complexes termed as "hot spots" (Clackson and Wells 1995; Ma et al. 2003; Moreira et al. 2007; Wang et al. 2018). Analysis of hot spots revealed that tryptophan (21%), arginine (13.1%), and tyrosine (12.3%) are most frequent and vital for the interaction. The hot spot is surrounded by the residues with a less energetic contribution for interaction that form O-ring-like shape so as to interrupt the entry of bulk solvent (Bogan and Torn 1998). Alanine scanning mutagenesis is a general approach for the identification of hot spot residues. In this technique interface residues are sequentially mutated to alanine and every time measured for binding affinity to associate protein. The residue which on mutation causes binding free energy difference ($\Delta\Delta G_{binding}$) $\geq$ 2.0 kcal/mol is identified as hot spot residue (Moreira et al. 2007). Several PPI analysis revealed that the structure of these hot spot regions on interfaces is highly conserved (Moreira et al. 2007; Wang et al.

2018), and their identification can add important steps towards our understanding of protein–protein interactions.

Aloy and Russell (2004), based on their assumption about homologous proteins, proposed that around 10,000 unique protein–protein interactions are possible. However, now it is well known that proteins from different families can form similar interface interactions and complex structures. Nowadays, the increasing information about hub proteins having many protein partners uncovered another fact that they use different surfaces or faces for interaction with distinct partners. It is interesting to know further about proteins using different faces for the same partner and proteins with the same faces for different partners. Kim et al. (2006) explained the term "faces" as a set of interface residues on a single domain contacting with another domain within 5 Å. So, one can define hot spots as the complemented pockets of structurally conserved residues scattered on the interfaces. Due to the complementarity of these pockets in terms of shape and juxtaposition of amino acid residues, the hot spot of one face packed with that of another partner (Moreira et al. 2007). The intense proclivity of hot spots towards different protein partners suggests the importance of its study not only for the understanding of protein–protein dimer but also for getting information about faces interacting with other molecules (Thornton 2001).

6.3 Wet-Lab Methods of PPI Detection (Table 6.1)

6.3.1 In Vitro PPI Modeling

6.3.1.1 Nuclear Magnetic Resonance (NMR)

During the twentieth century, NMR was considered as the best method for the study of three-dimensional structures of macromolecules. In recent times due to the development of better equipment, NMR can be used for the analysis of PPIs without the need for protein crystallization (Zuiderweg 2002). The crystallization of protein complexes in biologically relevant form might not be possible for every case, especially ones formed due to weak interactions, thus increase the importance of the NMR technique. However, it is a low-throughput method of protein–protein interaction study. Still, a range of NMR methods including chemical shift perturbation analysis, nuclear Overhauser effects (and its derivatives), residual dipolar couplings, paramagnetic approaches, solid-state NMR, and the analysis of low-abundance species can be used to understand the PPIs (Nishida and Shimada 2011).

Chemical shift perturbation is the most excepted NMR method to record protein interfaces. A two-dimensional NMR experiment called "Heteronuclear single quantum correlation" correlates the amide attached 15 N frequency (chemical shift) with the directly attached 1 H for every single amino acid. The interaction between proteins affects the chemical shifts of the nuclei in the protein interfaces (Vaynberg and Qin 2006). A full three-dimensional structure of protein complexes can be determined using the nuclear Overhauser effect (NOE) that measures inter-proton distances. However, this method is limited to the large size protein complexes

Table 6.1 A summarized overview of wet-lab methods for study of PPIs

Method	System	Index for PPI	Pros	Cons	References
Nuclear magnetic resonance (NMR)	In vitro system	Characteristic chemical shift	Rapid assay, can be applied in combination with in vivo approaches	Low throughput	Nishida and Shimada (2011), Sugiki et al. (2018)
Small angle X-ray scattering (SAXS)	In vitro system	Elastically scatter X-rays at small angles	Provide near-physiological conditions, applicable as integrative approaches	Challenge in binding of probes to targeted sites	Allec et al. (2015), Pons et al. (2010), Röllen et al. (2018)
Co-immunoprecipitation (Co-IP)	In vitro system (cell extracts)	Immunostaining	Provide native conditions for interactions study	Not appropriate for transient interactions	Moresco et al. (2010), Xing et al. (2016)
Cryo-electron microscopy (Cryo-EM)	In vitro system	Integration of 3D atomic model of protein partners	Crystallization of samples not required	Computer assisted data manipulation required	Milne et al. (2013), Costa et al. (2017)
Atomic force microscopy	In vitro system	Energy with respect to distance	Determines interaction forces between molecules at native physiological conditions	Highly sensitive to the sample preparation	Lin et al. (2005), Bangalore and Tessmer (2018)
Isothermal titration calorimetry	In vitro system	Change in thermodynamic properties	Quantitative method	Pure samples of proteins required	Pierce et al. (1999), Privalov and Dragan (2006)
Pull-down assays	In vitro system	Immunostaining	Rapid, quantitative approach	Protein tag might influence results	Pollard (2010), Louche et al. (2017)
Far-Western blotting	In vitro system	Radioactive detection/ immunostaining	Can check the effect of posttranslational modifications on PPIs	Pure proteins required, false-positive interactions	Machida and Mayer (2009), Hall (2015)
Tandem affinity purification-mass spectroscopy	In vitro system	Mass spectroscopy	High-throughput detection	Not suitable for labile and transient interactions	Podobnik et al. (2016), Struk et al. (2018)
Protein microarrays	In vitro system	Immuno/ fluorescence detection	High through multiplexed study PPIs	Lack of particular standards for data collection, analysis and validation	Hurst et al. (2009), Tomizaki et al. (2010)

(continued)

Table 6.1 (continued)

Method	System	Index for PPI	Pros	Cons	References
Surface plasmon resonance	In vitro technique	Surface plasmon resonance	Real-time monitoring, label-free, can resolve association and dissociation rate constant	Immobilization required, hard to quantify weak binding interactions	Ro et al. (2006), Nikolovska-Coleska (2015), Moscetti et al. (2017)
Two-hybrid system	In vivo (yeast, bacterial, plant, mammalian) system	Survival on depleted medium or enzymatic	Detect transient interactions	False-positive interactions	Fields and Song (1989), Ehlert et al. (2006), Mehla et al. (2017), Riegel et al. (2017)
Synthetic lethality	In vivo system	Variation in expected phenotype	Based on functional interactions	Not provide definite evidence	Ooi et al. (2006), Michaut and Bader (2012), Benstead-Hume et al. (2019)
Fluorescence resonance energy transfer (FRET)	In vivo system	Fluorescence	Real-time measurement of PPI	High background noise, reliance of signal on amount of fluorophores	Fernández-Dueñas et al. (2012), Cui et al. (2019)
Bioluminescence resonance energy transfer (BRET)	In vivo system	Bioluminescence	Detection can be performed in both cell culture and deep tissue	Cytotoxicity through bleaching	Pfleger and Eidne (2005), Dimri et al. (2016), Dale et al. (2019)
Luminescence-based mammalian interactome mapping (LUMIER)	Mammalian system	Bioluminescence	High-throughput semiquantitative technique, physiologically validate protein associations	Expensive technique requires high expertise	Vizoso Pinto et al. (2009), Blasche and Koegl (2013), Taipale (2018)
Mammalian protein–protein interaction trap (MAPPIT)	Mammalian system	Induced transcription of specific target gene	Provides extra level of control over false positives	High expertise required	Eyckerman et al. (2005), Uyttendaele et al. (2007), Vyncke et al. (2019)

(Kenworthy 2001; Vaynberg and Qin 2006). The accuracy of the NOE method can be improved, employing residual dipolar coupling (RDC) as an additional checkpoint. The relative orientation of certain dipoles existing within the complex in the magnetic field helps to determine the relative orientation of interacting subunits in PPIs (Prestegard et al. 2004).

Protein–ligand binding can be detected by paramagnetic NMR (paramagnetic relaxation enhancement, PRE), which increases the transverse relaxation rate due to dipolar interactions with unpaired electron immobilized on the protein (Sugiki et al. 2018). The molecular size and solubility limitation of NMR techniques have overcome by solid-state NMR (ssNMR), utilizing the approaches of magic angle sample spinning coupled with cross-polarization (Miao and Cross 2013). ssNMR has enabled us to characterize interfaces of membrane protein–ligand or other protein–protein complexes (Sun et al. 2012a). The NMR technique proceeds further with titration to get a proper estimation of the affinity, stoichiometry, specificity, and kinetics of binding or interaction between protein partners (Zuiderweg 2002).

6.3.1.2 Small Angle X-Ray Scattering (SAXS)

In contrast to the crystallography, which limited to the molecules with the ability to form crystals, small-angle X-ray scattering (SAXS) is a solution (Grant et al. 2011). SAXS is an effective technique to resolve structural features of biomolecules in the range between 1 and 100 nm at near-physiological conditions. In recent times, the SAXS method gets popular for the study of protein complexes (Allec et al. 2015; Schindler et al. 2016). SAXS was typically used for the detection of shape, size, distributions, and locations of various nanostructures, or can be used to reconstruct the low-resolution image of the molecular envelope. This method needs highly scattering molecular small size probes that bind with high affinity to selectively targeted biomolecules, not only surface but buried 3D structures without averting perturbing the system. The probes are designed in such a way that they elastically scatter X-rays at small angles above the background signal generated from the associated biomolecule (Koch et al. 2003). Protein–protein interaction can be screened using several different probes. Targeted biomolecules, when reaching in the zone of 1–100 nm proximity, produce signature scattering depending on the intermolecular distance and characteristic of interacting partners (Allec et al. 2015).

Grant et al. (2011) studied 28 different samples through crystallography, NMR, and SAXS methods and concluded that SAXS can enhance the structural information gathered by different techniques and is helpful in the interpretation of functional information from structural details. Several workers adopted an integrative approach with SAXS data to inspect protein–protein complexes and understand interactions on the genome-scale (Petoukhov and Svergun 2005; Hura et al. 2009; Pons et al. 2010; Schneidman-Duhovny et al. 2012; Karaca and Bonvin 2013; Xia et al. 2015). Even though binding of the probe to the targeted site of proteins is a challenge, SAXS utilizing optimally designed probes that generate sufficiently high scattering signals has the potential to be applied under in vitro, in cellulo, and in vivo conditions (Röllen et al. 2018).

6.3.1.3 Co-immunoprecipitation (Co-IP)

Co-immunoprecipitation is a very impressive technique as it utilizes cell extract containing proteins and cellular components providing native conditions to study interactions (Moresco et al. 2010). In this technique, a target antigen (so-called bait) is bound to an antibody (affinity tag) that is immobilized to a support material. The bait protein interacts with other protein(s) present in cell extract and gets precipitated due to the affinity tag attached to support. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis allow the detection of immunoprecipitated bait protein and interacting protein partners. The identification of peptides is possible through the excision of protein bands followed by mass spectrometry (Mann et al. 2001; Xing et al. 2016). This technique allows the study of weakly interacting proteins; however, a disadvantage is the requirement of highly specific tag antibodies. This difficulty can be resolved by using transfected cells expressing tagged bait protein. Now an antibody specific against tag (instead of for bait) can be utilized for the detection of interacting protein partners through co-IP (Masters 2004; Berggård et al. 2007).

6.3.1.4 Cryo-Electron Microscopy (Cryo-EM)

Day-by-day advancement in electron microscopy instrument and image processing system has made possible the study of protein complexes of sizes from hundreds of kilo-daltons to mega-daltons at 3–5 Å resolution (Costa et al. 2017). Crystallization of samples not required is an advantage of this technique. Cryo sample preparation is the sensational achievement, allowing the high resolution of EM structures providing the name cryo-EM to the technique. In this process, exposure of frozen solution of protein to an electron beam causes scattering of electrons that passes through the lens to create a magnified image (Callaway 2015). The frequently used variant of cryo-EM is single-particle cryo-EM, which combines a large number of 2D images of a protein complex in different orientations and creates a 3D image of the molecule. After the generation of such a 3D atomic model for the protein partners of complex the data integration into the density map to create pseudo-atomic models enhance the performance of electron microscopy (Milne et al. 2013). Success in the understanding of some Gram-negative bacterial type IV-secretion system assembly and functionally is a major output of cryo-EM (Costa et al. 2017).

6.3.1.5 Atomic Force Microscopy

As earlier discussed, the intermolecular forces operate and determine the protein interactions, so the method that will provide information about the distribution of interaction energy in space between two molecules is of great importance. Atomic force microscopy (AFM) makes available such information by measurement of force (energy with respect to distance) between two molecules (Leckband 2000; Lin et al. 2005). The important features of AFM are high force sensitivity of ~10-2pN (Lee et al. 1994), the dynamic range of ~0.001–5000 nN (Dammer et al. 1995), positional accuracy by 0.01 mn, and above all functional at native physiological conditions. The process involves scanning the molecule surface using a very sharp probe attached at the end of a cantilever, scrutinizing the deflection of the lever, and

reconstruction of a 3D topographical image. The materialistic properties (elasticity, electrostaticity, adhesion, and viscosity) of proteins can be measured at a spatial level due to the high force sensitivity of AFM (Chen et al. 1992; Clausen-Schaumann et al. 2000; Bangalore and Tessmer 2018). However, a very important limitation that is still associated with this technique is the sensitivity to the sample preparation and the condition of force measurement.

6.3.1.6 Isothermal Titration Calorimetry (ITC)

Protein–protein interaction causes structural changes in protein partners. An exclusive reorganization of bound water molecules due to interaction results in entropic and enthalpic changes in the system. Removal of water molecules from nonpolar surface is an enthalpically unfavorable but entropically favorable process. Balance of the enthalpy and entropy changes caused due to the type and number of bonds formed between protein interfaces develops a thermodynamic fingerprint, and that biophysics is the base of isothermal titration calorimetry (ITC) (Ladbury and Chowdhry 1996). This technique follows a direct and quantitative approach for the thermodynamic characterization of an interactive system (Doyle 1997; Pierce et al. 1999). ICT involves the incremental injection of one protein into a second protein sample taken in the calorimetric cell. Whether it is exothermic or endothermic reaction, measurement of heat of reaction is taken as a function of the concentration of protein complex formed during the binding reaction (Velazquez-Campoy et al. 2004). ITC is well suited for the study of heterodimeric interactions and dissociation of homodimers. It is also possible to quantify thermodynamic contributions of specific amino acids intercede the binding reaction on the availability of structural information of interfaces of interacting proteins (Privalov and Dragan 2006).

6.3.1.7 Pull-Down Assays

Pull-down assay is quantitative in vitro method to screen or verify the direct protein–protein interactions. The methodology of this assay is somewhat identical to co-immunoprecipitation in terms of the requirement of an affinity ligand to confine interacting proteins (Louche et al. 2017). A bait used in the pull-down method is a purified and tagged protein, while in case of co-immunoprecipitation is an immobilized antibody. This method uses immobilized tagged protein (bait) on tag-specific affinity ligand and prey protein in solution. To determine the binding affinity of bait and prey proteins, the bait at constant concentration allows interacting with the increasing concentrations of prey until binding sites get saturated. Now the bound prey-bait proteins complex eluted followed by the addition of competitive reagents, pH changes, or boiling of beads to disturb the interactions. The protein fraction is then resolved on sodium dodecyl sulfate-polyacrylamide gel followed by gel staining and scanning by densitometer to quantify band intensities. The dissociation constant (K-d) calculated by putting this data on a curve represents the binding affinity of bait-prey proteins (Pollard 2010). The specific sites which are necessary and sufficient for interaction can be screened using point mutated either bait or prey proteins or both of them (Lapetina and Gil-Henn 2017).

6.3.1.8 Far-Western Blotting

Far-Western blotting is a popular molecular method to detect in vitro protein–protein interactions. This method is quite similar to the traditional Western blotting, which involves separation of protein sample on SDS-PAGE or Native PAGE gel, then to transfer on the membrane. However, through Western blotting detection of protein is possible if the target is known in advance. In contrast to that far-Western blotting detects the binding sites with no need for the known identity of protein. In this sequence, the basic difference between these two methods is the use of antibodies as a probe in Western blotting while labeled "bait" proteins are used to probe target "prey" protein in case of the far-Western blot (Wu et al. 2007; Machida and Mayer 2009). Another important point of consideration is that the proteins get denatured during separation on SDS-PAGE gel, not suitable to detect protein–protein interactions and depends upon secondary and tertiary structures of proteins. A thoughtful implementation of the far-Western blotting method is the use of guanidine or urea for denaturation of proteins, and again renaturation and recovery of secondary/tertiary structures by gradual reduction of guanidine or urea (Hall 2015). Now "bait" protein probe is used for targeted binding to "prey" protein that observed using a detection system according to the labeled probe. By measurement of bands intensity, one can quantify the number and affinity between prey and bait proteins (Machida and Mayer 2009). Still, a limitation of this technique is the observation of false-positive results due to the development of non-native conformations following the denaturation-renaturation step that may interact with bait protein in a novel yet artificial way.

6.3.1.9 Tandem Affinity Purification-Mass Spectroscopy (TAP-MS)

Tandem affinity purification coupled with MS is a very effective method for high-throughput identification of protein complexes. In this method, the protein of interest is fused at C- or N-end with specially designed TAP tag by fusion of gene for a target protein with DNA fragment for tandem affinity protein (TAP) tag (Völkel et al. 2010; Podobnik et al. 2016). TAP tag is made up of two tags, the proximal calmodulin-binding peptide (CBP) and the distal IgG-binding segment of *Staphylococcus aureus* Protein A separated by either tobacco Etch virus (TEV) or rhinovirus 3C protease cleavage site (Rigaut et al. 1999; Van Leene et al. 2015). Under native conditions and endogenous promoters, the expression is endorsed followed by purification of binding protein partners with associated TAP-tag via a two-step affinity procedure. The first step involves the immobilization of protein complex on the IgG-containing matrix followed by specific elution through cleavage by TEV or rhinovirus protease. In a second step, the elution fraction from the former step is incubated with calmodulin-coated beads in the presence of Ca^{++} followed by washing for the removal of remains of protease, then elution of protein complex using chelating agents (Xu et al. 2010). These purification steps are taken over by SDS-PAGE separation of proteins from complexes, then excision from gel and digestion with trypsin. The amino acid sequence and mass data set generated by MS used to identify the protein and deduction of the PPI network. Though TAP-MS allows discrimination between specific protein–protein associations and nonspecific

protein background, because of a large number of purification steps it is not suitable for identification of labile and transient interactions (Pardo and Choudhary 2012; Struk et al. 2018).

6.3.1.10 Protein Microarrays

Protein microarrays or protein chips are an attractive high-throughput technology for multiplexed detection of protein–protein interactions that allows miniaturization and quantification (Kukar et al. 2002). The microarray-based study of PPIs is a vision before the year 2000 that materialized by MacBeath and Schreiber (2000). The first proteome based microarray was developed by Zhu et al. (2001) that contained 85% of the yeast genome and used to screen protein–protein interaction. Jones et al. (2006) demonstrated the quantification of PPIs on an array. However, several challenges associated with this technique include high-level expression of interacting proteins, their purification, surface immobilization of bait proteins keeping the tertiary structure intact (Hurst et al. 2009). Several bait-prey proteins detection strategies are in use, i.e., (i) use of prey-specific antibodies (Merbl and Kirschner 2009), (ii) expression of prey proteins associated with affinity tag like hemagglutinin (Ramachandran et al. 2004), (iii) single tag-specific antibody for multiple bait-prey associations (Hurst et al. 2009), (iv) fluorescent tag for small peptides and prey proteins for direct detection (Schnack et al. 2008), and (v) cell-free expression system for prey proteins for labeling with fluorophore, biotin, or 35*S*-methionine. With continuing improvements, this technique will become a leading tool for PPIs, especially for the diagnosis of diseases (Tomizaki et al. 2010).

6.3.1.11 Surface Plasmon Resonance

Surface plasmon resonance is a preferred method for the study or measurement of binding parameters of strong as well as weak strong protein–protein interactions in a range of mM to nM (Ohlson et al. 1997; Peess et al. 2015; Douzi 2017). This technique provides such strong optical data so that it can be used for the validation of interactions identified by other in vitro or in vivo methods (Nikolovska-Coleska 2015). In this method, sensor chip having carboxymethylated dextran monolayer attached to a gold surface used for immobilization of ligand molecules. The analyte solution now allows flowing over the surface of the immobilized ligand. Interaction of analyte to ligand alters mass concentration at the metal surface that consequently imitates in the refractive index, and thus generates resonance/response units (RUs) (Moscetti et al. 2017). SPR provides kinetic parameters of interaction between protein partners on a real-time basis even without labeling, and thus a well-suited tool for drug discovery and diagnosis of diseases (Mariani and Minunni 2014; Patching 2014). Another advantage of this technique is the ability to explore interactions between many proteins that communicate in combined, synergistic, or competitive modes (Ro et al. 2006).

6.3.2 In Vivo PPI Modeling

6.3.2.1 Two-Hybrid System

The high-throughput two-hybrid systems follow the same approach of restoration of protein activity through non-covalent binding between fragments of split protein. One fragment of a selected modular protein is fused to a protein of interest called "bait," and other fragment fused to "prey" protein. After co-expression, bait and prey proteins might interact and reconstitute the activity of modular protein factor that screened through the expression of the reporter gene (Stynen et al. 2012; Mehla et al. 2017). Foundation of these methods was developed by Stanley Fields and Ok-Kyu Song in 1989, using *Saccharomyces cerevisiae* transcription factor Gal4 as a modular protein for the production of the two-hybrid system, the so-called "yeast two-hybrid system" (Fields and Song 1989). Two hybrids are developed between proteins of interest and either the DNA-binding domain or the activation domain of transcription factor Gal4. Yeast two-hybrid system can detect transient interactions, but has constraint due to false-positive rate and limited information generation about kinetics and dynamics of PPIs. Proteins that under natural conditions reside in different subcellular locations might detect as interacting partners by this method. The attempt to overcome this problem leads to a reduction in the efficiency of this method (Brückner et al. 2009).

The bacterial two-hybrid system is another choice that is more preferably used for bacterial proteins. In the initial 1990s, a bacterial two-hybrid system was developed using dimeric λ repressor protein of *Escherichia coli* (Hu et al. 1990; Edgerton and Jones 1992). Replacement of the C-terminal domain of λ repressor protein with the protein of interest inhibits the binding of its N-terminal domain on DNA, which is detected as repression of reporter gene lacZ. Bacterial adenylate cyclase based two-hybrid system is another scheme of current interest. Joining of 25 kDa catalytic site (T25) to 18 kDa calmodulin-binding site (T18) is necessary for the activation of adenylate cyclase and synthesis of cAMP. These two T25 and T18 domains of adenylate cyclase come in close immediacy on the interaction of bait and prey proteins and lead the production of cAMP. Binding of cAMP to a catabolic activator protein induce transcription of catabolic operons of lactose or maltose. The interaction of bait and prey proteins is detected in the form of the characteristic phenotype of *E. coli* on indicator media (Mehla et al. 2017).

Interactions between proteins of plant or mammalian origin detected through yeast or bacterial two-hybrid system might provide the wrong idea due to differences in the cellular system. A plant two-hybrid system using *Arabidopsis thaliana* protoplast was developed by Ehlert et al. (2006). Basic leucine zipper (bZIP) transcription factors having characteristics of homo- and heterodimerization were selected as a system for the detection of protein–protein interactions. On comparison of yeast and plant two-hybrid systems, Ehlert et al. (2006) found that weak heterodimerization actions were detected through plant two-hybrid system, however, remained undetected in a yeast system. In a similar way, use of mammalian two-hybrid systems that allows posttranslational changes is desirable for the study of interactions between mammalian proteins. Though the mammalian system offers

perfect physiological conditions, it still does not provide high sensitivity for the quantitative detection of mammalian protein–protein interactions (Riegel et al. 2017). Nowadays several variants of hybrid system are available, viz. reverse two-hybrid systems, three-hybrid systems, two-bait hybrid systems, etc., that can be of choice according to the type/origin of proteins for the studies of the interaction (Ehlert et al. 2006; Brückner et al. 2009; Stynen et al. 2012; Riegel et al. 2017).

6.3.2.2 Synthetic Lethality

Synthetic lethality expresses a connection between a gene pair in which mutation in the individual gene is nonlethal; however, mutation of both genes leads to cell death (Bender and Pringle 1991; Ye et al. 2005). Proteins encoded by an enormous number of synthetic lethal gene pairs share interaction partners and perform interrelated functions (Appling 1999; Talavera et al. 2013). Protein–protein interaction is the fundamental base of synthetic lethality. Consider a mutation in gene "a" of gene pair "ab" weakens the interaction of its protein product "A" with "B." Another mutation now this time in gene "b" further weakens the interaction between "A" and "B" to a synthetically lethal level. SSL interactions represent a type of negative genetic interaction characterized by variation in expected phenotype due to mutations in more than one gene (Michaut and Bader 2012).

In this method, mutations are generated in genes assumed to be the part of a gene pair and observed for lethality only in case when both genes are mutated (Ooi et al. 2006; Rao et al. 2014). This method indirectly detects the interaction between proteins, referred to as functional interactions. Nguyen et al. (1998) identified various components of a g-tubulin complex of yeast by synthetic lethality method and further verified this functional interaction through two-hybrid systems. In recent years synthetic lethality appears as an important tool that allows identification of specific targets of anticancer drugs (Liu et al. 2018; Benstead-Hume et al. 2019). However, this method does not provide definitive evidence for protein–protein interaction of any particular pathway. But the indication provided by synthetic lethality method can be used to trace the role of that particular protein–protein interaction in a pathway through subsequent biochemical or cell biological experiments.

6.3.2.3 Fluorescence Resonance Energy Transfer (FRET)

The principle behind resonance energy transfer is the post-excitation redistribution of electron energy either between or within molecules, first explained by Förster (1948). To study interactions among molecules, a technique developed based on this principle is known as Förster or fluorescence resonance energy transfer (FRET). The technique involves dipole-dipole energy transfer from fluorescent donor to acceptor with the condition that the distance between two fluorophores (donor and acceptor) must be less than 10 nm (Cui et al. 2019). To do so, target proteins attached to fluorophores either through fluorophore-tagged antibodies or by direct fusion. Fluorescence intensity ratio is depicted as a decrease in fluorescence intensity of donor fluorophore to that of the simultaneous increase in the intensity of acceptor fluorophore, which is used as a measure of FRET.

Several methods are available to measure FRET, viz. wide-field standard epi-fluorescence microscopy, multi-photon, and confocal laser scanning microscopy (Parsons et al. 2004; Jares-Erijman and Jovin 2006; Margineanu et al. 2016). High background noise, insensitivity, and dependence of the signal on the amount of fluorophores are some limitations of this method that can be resolved to a certain extent using fluorescence lifetime imaging microscopy (FLIM), which collects data of fluorescence lifetime of donor fluorophore (Sun et al. 2012b). Rainey and Patterson (2019) developed a method called photoswitching FRET that offers a non-complicated approach for improvement in FRET imaging. Meiresonne et al. (2018) studied the protein interactions in the cytoplasm as well as in periplasm of *E. coli* utilizing the FRET method based on spectroscopy. FRET-based methods allow experimentation in real-time providing comprehensive and accurate knowledge about protein–protein interactions (Fernández-Dueñas et al. 2012). The scope of FRET methods for the study of molecular interaction will increase with continuing innovations in fluorescent probes, microscopy, and image quantification systems.

6.3.2.4 Bioluminescence Resonance Energy Transfer (BRET)

BRET was first observed in sea pansy *Renilla reniformis* and jellyfish *Aequoria victoria*. BRET technique is based on the principle of resonance transfer and appeared as a striking substitute to a FRET method for the real-time study of PPIs in living cells, cell lysates, or purified proteins (Xu et al. 1999). In this method, the cDNAs for interacting proteins were engineered to express on protein fused to the luminescent donor (Renilla Luciferase) and other partners associated with fluorescent acceptor protein (usually a yellow fluorescent protein, YFP) (Dimri et al. 2016). Now, coelenterazine (CLZN), a cell-permeable substrate, is added so that it can undergo catalytic degradation on the interaction between proteins and the released energy transferred for luciferase to YFP. Luminescent signals from luciferase are checked at 480 nm, while light emitted by TFP is measured at 530 nm. Less than 10 nm distance between interacting partners is required for resonance energy transfer, a feature perfect for the study of PPIs (Harikumar et al. 2017; Dale et al. 2019).

Nowadays, several approaches and new small modified nanoluciferase (NanoLuc) are developed that are generally a 150-fold higher signal with four times more stable signal half-life over Renilla Luciferase (Hall et al. 2012; El Khamlichi et al. 2019). In several aspects, BRET is superior over FRET because fluorescence based technique has drawbacks of photobleaching, autofluorescence, and simultaneous excitation of donor and acceptor fluorophores. However, BRET also has a disadvantage of the requirement of donor protein as a part of the fusion protein, while in the case of FRET donor and acceptor fluorophores are conjugated to antibodies, thus a better choice for the study of cell surface interactions (Pfleger and Eidne 2005, 2006).

6.3.2.5 Luminescence-Based Mammalian Interactome Mapping

To detect protein–protein interactions in mammalian cell systems, luminescence-based mammalian interactome mapping technique was developed by Barrios-Rodiles et al. (2005). This is a semiquantitative technique that allows the detection or mapping of signaling pathways dependent on posttranslational modifications (Deng et al. 2014; Taipale 2018). The LUMIER technique involves co-expression of luciferase-tagged bait and N-terminal Flag-tagged prey proteins in an efficient cell line through transfection. To detect the interacting proteins in an extract from the cell lysate are enriched for Flag-tagged prey protein via affinity purification using sepharose beads and Flag-specific antibodies. The next step is the detection of generated bioluminescence on the addition of a substrate for luciferase as an indicator of bound bait protein in enriched extract (Blasche and Koegl 2013). Various luciferases can be employed for this purpose, such as Renilla luciferase (RLUC or RL, 36 kDa) and Firefly luciferase (64 kDa) (Barrios-Rodiles et al. 2017).

In a modified method, luminescence-based maltose-binding protein pull-down interaction screening system (LuMPIS), Vizoso Pinto and coworkers used maltose-binding protein (MBP) to tag prey protein and enhanced green fluorescence protein-luciferase (eGFP-Luc) tagged bait protein (Vizoso Pinto et al. 2009). An improvement in the technique is the use of genetically modified luciferase of much smaller size like NanoLuc (19 kDa) with high sensitivity for the optimized substrate. Incorporation of such small and modified luciferase allows increased throughput detection PPIs with luminometers without the utilization of ultra-sensitive photomultipliers, an advantage for network biology or cell signaling studies (Hall et al. 2012).

6.3.2.6 Mammalian Protein–Protein Interaction Trap

Mammalian protein–protein interaction trap (MAPPIT) is a novel high-throughput two-hybrid mammalian system that allows the detection of proteins interactions in cell physiological environment, even for those proteins that require posttranslational modifications. This method is based on the concept of type I cytokine receptor signal transduction (Uyttendaele et al. 2007). Binding of ligand to cytokine receptor elicits the cross-activation of linked Janus kinase (JAK), which eventually leads to the phosphorylation of tyrosine residue on the specific position of receptor tails. The phosphorylated position now turns into the docking site for signal transducer and activator of transcription (STAT) signaling molecules and subsequently activates recruit STAT, which moves to the nucleus and induces transcription of the specific target gene (Eyckerman et al. 2001; Levy and Darnell 2002; Tavernier et al. 2002).

Through the MAPPIT method, it is possible to even screen unidentified protein partners to study cDNA libraries following the above-discussed concept. For the purpose, specific HEK293T-derived cell lines can be developed by co-transfection and stable expression of bait, prey cDNA, and reporter gene (Vyncke et al. 2019). Bait is expressed as a fusion protein with STAT recruitment-deficient receptor but still able to activate JAK and prey protein fused to a receptor fragment containing functional STAT recruitment sites. Interaction of bait and prey proteins restores the activity of STAT so that the clones expressing interacting protein partners are

detected through the expression of stably integrated reporters such as luciferase (Ulrichts et al. 2009) or puromycin resistance marker (Eyckerman et al. 2002) under the control of STAT-responsive rPAP1 promoter. Now, the MAPPIT technique based reverse two-hybrid system (reverse MAPPIT) is also available for relatively easy detection and analysis of disruptor molecules (Eyckerman et al. 2005).

6.4 Dry-Lab (In Silico) Methods of PPI Detection

The function of a protein can be explained using the PPI relationship. Phenotype is not the result of a single isolated protein or gene, but it is the result of the interaction of many genes/proteins. The in vitro and in vivo methods of PPI have their advantages and limitations. There are different in silico approaches for PPI analysis, which are based on information derived from sequence, structure, chromosome proximity, in silico 2 hybrid, gene fusion, phylogenetic profile, and gene expression profile (Berggård et al. 2007). In most cases, proteins perform their function as a complex with other proteins. Over 80% of proteins do not operate as isolated species but in complexes with other. Proteins involved in the same cellular process interact with each other.

6.4.1 Sequence-Based Approaches

In this approach, PPI analysis is based on sequential homology. It is considered that a PPI relationship in one species can infer the interaction in other species if it shares significant sequence homology (Lee et al. 2008). Sequence-based prediction is two types: (1) ortholog-based approach and (2) domain-pairs-based approach. The ortholog-based approach infers the interaction based on the similarity of the target protein with a well-annotated sequence homolog. Many proteins in one target organism share significant similarities with proteins involved in a pathway in other organisms. The domain-pairs-based approach is based on conserved domains search, which is used for predicting structural class, subcellular location, enzyme class and subclass, and other predictions (Wojcik and Schächter 2001). Domains are involved in the intermolecular interaction and may be used to interpret PPI.

6.4.2 Structure-Based Approaches

PPI can be inferred based on the similarity between the two proteins. If the interaction between two proteins is known, then two other proteins that show significant similarity with these interacting proteins can also interact with each other (Zhang et al. 2012). The limitation of this method is that the structure of many proteins has not been determined yet.

6.4.3 Chromosome Proximity/Gene Neighborhood

This approach of PPI prediction assumes that the interacting proteins tend to closely lie into the genomes in prokaryotes. Functional linkage existing between proteins that are derived from the related genes can be compared. In adjacent bidirectionally transcribed genes, one gene encodes for a transcriptional regulator, and the other gene encodes for a non-regulatory protein (Yamada et al. 2003). This approach is only applicable for bacteria since gene neighboring is conserved in the bacteria.

6.4.4 Gene Fusion

Single-domain proteins in one organism can fuse to form a multidomain protein in other organisms (Marcotte et al. 1999). Domain fusion is seen in those proteins involved in the metabolic pathway. PPI analysis can be deduced by using the information of domain fusion in different genomes.

6.4.5 In Silico 2 Hybrid

The method is based on the assumption that the two interacting proteins undergo coevolution to conserve the function. If some amino acids underwent certain changes in one protein, then the related residues in the interacting protein should also make mutations to store the function (Pazos and Valencia 2002). Here, PPI is interpreted based on the correlation between the interacting protein and the individual proteins.

6.4.6 Phylogenetic Profile

During evolution, functionally interacting proteins or genes coexists and evolve together. If two proteins have a functional linkage, then they will be inherited together during evolution (Lin et al. 2013). A phylogenetic profile represents the presence of certain proteins or genes in a set of the organism. Two proteins may have a functional linkage if they share the same functional profile.

6.4.7 Gene Expression-Based Approaches

The expression of genes can be grouped into a certain cluster based on their expression levels. The functional association of the various genes can be explained using the expression level of genes under several experimental conditions. Genes that show common expression-profile have more chances to interact with each other than genes with different expression-profile (Grigoriev 2001). Gene co-expression is

not an appropriate way to infer PPI but can be used to validate the PPI results generated from other experimental methods.

6.5 Conclusions

Vast array of methods discussed here provides comprehensive information for the selection of technique to understand the molecular dynamics of protein–protein interaction. Regardless of latest developments in these techniques still have some limitations. It is necessary to consider these limitations in addition to the benefits to make a choice of technique for understanding the variety of PPIs. Therefore the data set from each technique must validate using appropriate methods. Alternative in vivo and in vitro techniques can take vital part in this data verification.

References

Allec N, Choi M, Yesupriya N, Szychowski B, White MR, Kann MG, Badano A (2015) Small-angle X-ray scattering method to characterize molecular interactions: proof of concept. Sci Rep 5(1):12085. https://doi.org/10.1038/srep12085

Aloy P, Russell RB (2004) Ten thousand interactions for the molecular biologist. Nat Biotechnol 22:1317–1321

Ansari S, Helms V (2005) Statistical analysis of predominantly transient protein-protein interfaces. Proteins 61:344–355

Appling DR (1999) Genetic approaches to the study of protein–protein interactions. Methods 19 (2):338–349

Bangalore DM, Tessmer I (2018) Unique insight into protein-DNA interactions from single molecule atomic force microscopy. AIMS Biophys 5(3):194–216. https://doi.org/10.3934/biophy.2018.3.194

Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, Luga V, Przulj N, Robinson M, Suzuki H, Hayashizaki Y, Jurisica I, Wrana JL (2005) High-throughput mapping of a dynamic signaling network in mammalian cells. Science 307(5715):1621–1625

Barrios-Rodiles M, Ellis JD, Blencowe BJ, Wrana JL (2017) LUMIER: a discovery tool for mammalian protein interaction networks. Proteomics 1550:137–148

Bender A, Pringle JR (1991) Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in *Saccharomyces cerevisiae*. Mol Cell Biol 11:1295–1305

Benstead-Hume G, Chen X, Hopkins SR, Lane KA, Downs JA, Pearl FMG (2019) Predicting synthetic lethal interactions using conserved patterns in protein interaction networks. PLoS Comput Biol 15(4):e1006888

Berggård T, Linse S, James P (2007) Methods for the detection and analysis of protein–protein interactions. Proteomics 7:2833–2842

Blasche S, Koegl M (2013) Analysis of protein–protein interactions using LUMIER assays. Methods Mol Biol 1064:17–27

Bogan AA, Torn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9

Braun P, Gingras AC (2012) History of protein-protein interactions: from egg-white to complex networks. Proteomics 12(10):1478–1498

Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci 10(6):2763–2788

Callaway E (2015) The revolution will not be crystallized: a new method sweeps through structural biology. Nature 525:172–174

Chen CH, Vesecky SM, Gewirth AA (1992) In situ atomic force microscopy of underpotential deposition of silver on gold(111). J Am Chem Soc 114:451–458

Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267:383–386

Clausen-Schaumann H, Seitz M, Krautbauer R, Gaub HE (2000) Force spectroscopy with single bio-molecules. Curr Opin Chem Biol 4:524–530

Costa TRD, Ignatiou A, Orlova EV (2017) Structural analysis of protein complexes by cryo electron microscopy. In: Journet L, Cascales E (eds) Bacterial protein secretion systems, Methods in molecular biology, vol 1615. Humana Press, New York, NY

Cui Y, Zhang X, Yu M, Zhu Y, Xing J, Lin J (2019) Techniques for detecting protein-protein interactions in living cells: principles, limitations, and recent progress. Sci China Life Sci 62 (5):619–632. https://doi.org/10.1007/s11427-018-9500-7

Dale NC, Johnstone EKM, White CW, Pfleger KDG (2019) NanoBRET: the bright future of proximity-based assays. Front Bioeng Biotechnol 7:56

Dammer U, Popescu O, Wagner P, Anselmetti D, Guntherodt HJ, Misevic GN (1995) Binding strength between cell adhesion proteoglycans measured by atomic force microscopy. Science 267:1173–1175

De S, Krishnadev O, Srinivasan N, Rekha N (2005) Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. BMC Struct Biol 5:15. https://doi.org/10.1186/1472-6807-5-15

Deng Q, Wang D, Li F (2014) Detection of viral protein-protein interaction by microplate-format luminescence-based mammalian interactome mapping (LUMIER). Virol Sin 29(3):189–192. https://doi.org/10.1007/s12250-014-3436-8

Dimri S, Basu S, De A (2016) Use of BRET to study protein–protein interactions *in vitro* and *in vivo*. Methods Mol Biol 1443:57–78

Douzi B (2017) Protein–protein interactions: surface plasmon resonance. Methods Mol Biol 1615:257–275

Doyle ML (1997) Characterization of binding interactions by isothermal titration calorimetry. Curr Opin Biotechnol 8:31–35

Edgerton MD, Jones AM (1992) Localization of protein-protein interactions between subunits of phytochrome. Plant Cell 4:161–171

Ehlert A, Weltmeier F, Wang X, Mayer CS, Smeekens S, Vicente-Carbajosa J, Dröge-Laser W (2006) Two-hybrid protein-protein interaction analysis in Arabidopsis protoplasts: establishment of a heterodimerization map of group C and group S bZIP transcription factors. Plant J 46 (5):890–900

El Khamlichi C, Reverchon-Assadi F, Hervouet-Coste N, Blot L, Reiter E, Morisset-Lopez S (2019) Bioluminescence resonance energy transfer as a method to study protein-protein interactions: application to G protein coupled receptor biology. Molecules 24(3):537

Eyckerman S, Lemmens I, Catteeuw D, Verhee A, Vandekerckhove J, Lievens S, Tavernier J (2005) Reverse MAPPIT: screening for protein–protein interaction modifiers in mammalian cells. Nat Methods 2:427–433

Eyckerman S, Lemmens I, Lievens S, Van der Heyden J, Verhee A, Vandekerckhove J, Tavernier J (2002) Design and use of a mammalian protein-protein interaction trap (MAPPIT). Sci STKE 162:pl18

Eyckerman S, Verhee A, Van der Heyden J, Lemmens I, Van Ostade X, Vandekerckhove J, Tavernier J (2001) Design and application of a cytokine-receptorbased interaction trap. Nat Cell Biol 3:1114–1119

Fernández-Dueñas V, Llorente J, Gandía J, Borroto-Escuela DO, Agnati LF, Tasca CI, Ciruela F (2012) Fluorescence resonance energy transfer-based technologies in the study of protein–protein interactions at the cell surface. Methods 57(4):467–472

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature 340:245–246

Förster T (1948) Intermolecular energy migration and fluorescence. Ann Phys 2:55–75

Geva G, Sharan R (2010) Identification of protein complexes from co-immunoprecipitation data. Bioinformatics 27(1):111–117

Grant TD, Luft JR, Wolfley JR, Tsuruta H, Martel A, Montelione GT, Snell EH (2011) Small angle X-ray scattering as a complementary tool for high-throughput structural studies. Biopolymers 95 (8):517–530

Grigoriev A (2001) A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. Nucleic Acids Res 29(17):3513–3519

Hall MP, Unch J, Binkowski BF, Valley MP, Butler BL, Wood MG, Otto P, Zimmerman K, Vidugiris G, Machleidt T, Robers MB, Benink HA, Eggers CT, Slater MR, Meisenheimer PL, Klaubert DH, Fan F, Encell LP, Wood KV (2012) Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. ACS Chem Biol 7(11):1848–1857

Hall RA (2015) Studying protein-protein interactions via blot overlay/Far Western blot. Methods Mol Biol 1278:371–379

Harikumar KG, Yan Y, Xu T, Melcher K, Xu HE, Miller LJ (2017) Bioluminescence resonance energy transfer (BRET) assay for determination of molecular interactions in living cells. Bio Protoc 7(22):e2904

Hu JC, O'Shea EK, Kim PS, Sauer RT (1990) Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. Science 250:1400–1403

Hura GL, Menon AL, Hammel M, Rambo RP, Poole Ii FL, Tsutakawa SE, Jenney FE, Jr Classen S, Frankel KA, Hopkins RC et al (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods 6:606–612

Hurst R, Hook B, Slater MR, Hartnett J, Storts DR, Nath N (2009) Protein–protein interaction studies on protein arrays: effect of detection strategies on signal-to-background ratios. Anal Biochem 392(1):45–53

Janin J, Bahadur RP, Chakrabarti P (2008) Protein-protein interaction and quaternary structure. Q Rev Biophys 41:133–180

Jares-Erijman EA, Jovin TM (2006) Imaging molecular interactions in living cells by FRET microscopy. Curr Opin Chem Biol 10:409–416

Jones RB, Gordus A, Krall JA, Macbeath G (2006) A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 439:168–174

Karaca E, Bonvin AM (2013) On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. Acta Crystallogr D Biol Crystallogr 69:683–694

Kenworthy AK (2001) Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. Methods 24:289–296

Kim WK, Henschel A, Winter C, Schroeder M (2006) The many faces of protein–protein interactions: a compendium of interface geometry. PLoS Comput Biol 2(9):e124

Koch MHJ, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36 (2):147–227

Kukar T, Eckenrode S, Gu Y, Lian W, Megginson M, She J-X, Wu D (2002) Protein microarrays to detect protein–protein interactions using red and green fluorescent proteins. Anal Biochem 306 (1):50–54

Ladbury JE, Chowdhry BZ (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. Chem Biol 3:791–801

Lapetina S, Gil-Henn H (2017) A guide to simple, direct, and quantitative in vitro binding assays. J Biol Methods 4(1):e62

Leckband D (2000) Measuring the forces that control protein interactions. Annu Rev Biophys Biomol Struct 29:1–26

Lee GU, Kidwell DA, Colton RJ (1994) Sensing discrete streptavidin-biotin interactions with atomic force microscopy. Langmuir 10:354–357
Lee SA, Chan CH, Tsai CH, Lai JM, Wang FS, Kao CY, Huang CY (2008) Ortholog-based protein-protein interaction prediction and its application to inter-species interactions. BMC Bioinformatics 9:S11
Levy DE, Darnell JE (2002) Stats: transcriptional control and biological impact. Nat Rev Mol Cell Biol 3:651–662
Lin S, Chen J-L, Huang L-S, Lin H-W (2005) Measurements of the forces in protein interactions with atomic force microscopy. Curr Proteomics 2(1):55–81
Lin TW, Wu JW, Chang DT (2013) Combining phylogenetic profiling-based and machine learning-based techniques to predict functional related proteins. PLoS One 8(9):e75940
Liu L, Chen X, Hu C, Zhang D, Shao Z, Jin Q, Ke K (2018) Synthetic lethality-based identification of targets for anticancer drugs in the human signaling network. Sci Rep 8(1):8440. https://doi.org/10.1038/s41598-018-26783-w
Louche A, Salcedo SP, Bigot S (2017) Protein–protein interactions: pull-down assays. Methods Mol Biol 1615:247–255
Ma B, Elkayam T, Wolfson H, Nussinov R (2003) Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc Natl Acad Sci U S A 100:5772–5777
MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. Science 289:1760–1763
Machida K, Mayer BJ (2009) Detection of protein-protein interactions by far-western blotting. Nat Protoc 2:313–329
Mann M, Hendrickson RC, Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem 70:437–473
Marcotte EM, Pellegrini M, Ng H-L, Rice DW, Yeates TO, Eisenberg D (1999) Detecting protein function and protein-protein interactions from genome sequences. Science 285(5428):751–753
Margineanu A, Chan JJ, Kelly DJ, Warren SC, Flatters D, Kumar S, French PMW (2016) Screening for protein-protein interactions using Förster resonance energy transfer (FRET) and fluorescence lifetime imaging microscopy (FLIM). Sci Rep 6(1):28186. https://doi.org/10.1038/srep28186
Mariani S, Minunni M (2014) Surface plasmon resonance applications in clinical analysis. Anal Bioanal Chem 406:2303–2323
Masters SC (2004) Co-immunoprecipitation from transfected cells. Methods Mol Biol 261:337–350
Mehla J, Caufield JH, Sakhawalkar N, Uetz P (2017) A comparison of two-hybrid approaches for detecting protein–protein interactions. Methods Enzymol 586:333–358. https://doi.org/10.1016/bs.mie.2016.10.020
Meiresonne NY, Alexeeva S, van der Ploeg R, den Blaauwen T (2018) Detection of protein interactions in the cytoplasm and periplasm of *Escherichia coli* by förster resonance energy transfer. Bio Protocol 8(2):e2697
Merbl Y, Kirschner MW (2009) Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. Proc Natl Acad Sci U S A 2009(106):2543–2548
Miao Y, Cross TA (2013) Solid state NMR and protein-protein interactions in membranes. Curr Opin Struct Biol 23(6):919–928. https://doi.org/10.1016/j.sbi.2013.08.004
Michaut M, Bader GD (2012) Multiple genetic interaction experiments provide complementary information useful for gene function prediction. PLoS Comput Biol 8:e1002559. https://doi.org/10.1371/journal.pcbi.1002559
Milne JL, Borgnia MJ, Bartesaghi A, Tran EE, Earl LA, Schauder DM, Lengyel J, Pierson J, Patwardhan A, Subramaniam S (2013) Cryo-electron microscopy—a primer for the non-microscopist. FEBS J 280:28–45
Mintseris J, Weng Z (2005) Structure, function, and evolution of transient and obligate protein–protein interactions. PNAS 102(31):10930–10935

Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots—a review of the protein-protein interface determinant amino-acid residues. Proteins 68(4):803–812

Moresco JJ, Carvalho PC, Yates JR III (2010) Identifying components of protein complexes in C. elegans using co-immunoprecipitation and mass spectrometry. J Proteome 73 (11):2198–2204

Moscetti I, Cannistraro S, Bizzarri AR (2017) Surface plasmon resonance sensing of biorecognition interactions within the tumor suppressor p53 network. Sensors 17:2680. https://doi.org/10.3390/s17112680

Nguyen T, Vinh DBN, Crawford DK, Davis TN (1998) A genetic analysis of interactions with Spc110p reveals distinct functions of Spc97p and Spc98p, components of the yeast gamma-tubulin complex. Mol Biol Cell 9:2201–2216

Nikolovska-Coleska Z (2015) Studying protein-protein interactions using surface plasmon resonance. Methods Mol Biol 1278:109–138

Nishida N, Shimada I (2011) An NMR method to study protein–protein interactions. In: Shimaoka M (ed) Integrin and cell adhesion molecules, Methods in molecular biology (methods and protocols). Humana Press, Totowa, NJ, p 757

Nooren IMA, Thornton JM (2003) Diversity of protein-protein interactions. EMBO J 22 (14):3486–3492

Ohlson S, Strandh M, Nilshans H (1997) Detection and characterization of weak affinity antibody antigen recognition with biomolecular interaction analysis. J Mol Recognit 10:135–138

Ooi SL, Pan X, Peyser BD, Ye P, Meluh PB, Yuan DS, Irizarry RA, Bader JS, Spencer FA, Boeke JD (2006) Global synthetic-lethality analysis and yeast functional profiling. Trends Genet 22 (1):56–63

Pardo M, Choudhary JS (2012) Assignment of protein interactions from affinity purification/mass spectrometry data. J Proteome Res 11:1462–1474

Park SH, Reyes JA, Gilbert DR, Kim JW, Kim S (2009) Prediction of protein-protein interaction types using association rule based classification. BMC Bioinformatics 10:36

Parsons M, Vojnovic B, Ameer-Beg S (2004) Imaging protein–protein interactions in cell motility using fluorescence resonance energy transfer (FRET). Biochem Soc Trans 32(3):431–433. https://doi.org/10.1042/bst0320431

Patching SG (2014) Surface plasmon resonance spectroscopy for characterisation of membrane protein—ligand interactions and its potential for drug discovery. Biochim Biophys Acta 1838:43–55

Pazos F, Valencia A (2002) In silico two-hybrid system for the selection of physically interacting protein pairs. Proteins 47(2):219–227

Peess C, von Proff L, Goller S, Andersson K, Gerg M, Malmqvist M, Bossenmaier B, Schräml M (2015) Deciphering the stepwise binding mode of HRG1beta to HER3 by surface plasmon resonance and interaction map. PLoS One 10:e0116870

Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C (2010) Transient protein-protein interactions: structural, functional, and network properties. Structure 18(10):1233–1243

Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89:1237–1250

Pfleger KDG, Eidne KA (2005) Monitoring the formation of dynamic G protein-coupled receptor-protein complexes in living cells. Biochem J 385:625–637

Pfleger KDG, Eidne KA (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). Nat Methods 3(3):165–174

Pierce MM, Raman CS, Nall BT (1999) Isothermal titration calorimetry of protein-protein interactions. Methods 19:213–221

Podobnik M, Kraševec N, Zavec AB, Naneh O, Flašker A, Caserman S, Hodnik V, Anderluh G (2016) How to study protein-protein interactions. Acta Chim Slov 63:424–439

Pollard TD (2010) A guide to simple and informative binding assays. Mol Biol Cell 21 (23):4061–4067

Pons C, D'Abramo M, Svergun DI, Orozco M, Bernadó P, Fernández-Recio J (2010) Structural characterization of protein–protein complexes by integrating computational docking with small-angle scattering data. J Mol Biol 403:217–230

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104(8):3519–3540

Privalov PL, Dragan AI (2006) Microcalorimetry of biological macromolecules. Biophys Chem 122:158–169

Rainey KH, Patterson GH (2019) Photoswitching FRET to monitor protein-protein interactions. Proc Natl Acad Sci U S A 116(3):864–873

Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, Lau AY, Walter JC, LaBaer J (2004) Self-assembling protein microarrays. Science 305:86–90

Rao VS, Srinivas K, Sujini GN, Kumar GNS (2014) Protein-protein interaction detection: methods and analysis. Int J Proteomics 2014:1–12. https://doi.org/10.1155/2014/147648

Riegel E, Heimbucher T, Höfer T, Czerny T (2017) A sensitive, semi-quantitative mammalian two-hybrid assay. Biotechniques 62(5):206–214. https://doi.org/10.2144/000114544

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 17:1030–1032

Ro H-S, Koh BH, Jung SO, Park HK, Shin YB, Kim M-G, Chung BH (2006) Surface plasmon resonance imaging protein arrays for analysis of triple protein interactions of HPV, E6, E6AP, and p53. Proteomics 6:2108–2111

Röllen K, Granzin J, Batra-Safferling R, Stadler AM (2018) Small-angle X-ray scattering study of the kinetics of light-dark transition in a LOV protein. PLoS One 13(7):e0200746

Sali A, Glaeser R, Earnest T, Baumeister W (2003) From words to literature in structural proteomics. Nature 422:216–225

Schindler CEM, de Vries SJ, Sasse A, Zacharias M (2016) SAXS data alone can generate high-quality models of protein-protein complexes. Structure 24(8):1387–1397

Schnack C, Danzer KM, Hengerer B, Gillardon F (2008) Protein array analysis of oligomerization-induced changes in a-synuclein protein–protein interactions points to an interference with Cdc42 effector proteins. Neuroscience 154:1450–1457

Schneidman-Duhovny D, Rossi A, Avila-Sakar A, Kim SJ, Velázquez Muriel J, Strop P, Liang H, Krukenberg KA, Liao M, Kim HM et al (2012) A method for integrative structure determination of protein-protein complexes. Bioinformatics 28:3282–3289

Selzer T, Schreiber G (2001) New insights into the mechanism of protein-protein association. Proteins 45(3):190–198. https://doi.org/10.1002/prot.1139

Struk S, Braem L, Walton A, De Keyser A, Boyer F-D, Persiau G, De Jaeger G, Gevaert K, Goormachtig S (2018) Quantitative tandem affinity purification, an effective tool to investigate protein complex composition in plant hormone signaling: strigolactones in the spotlight. Front Plant Sci 9:528

Stynen B, Tournu H, Tavernier J, Van Dijck P (2012) Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. Microbiol Mol Biol Rev 76(2):331–382

Sugiki T, Furuita K, Fujiwara T, Kojima C (2018) Current NMR techniques for structure-based drug discovery. Molecules 23:148. https://doi.org/10.3390/molecules23010148

Sun S, Paramasivam S, Yan S, Williams JC, Byeon In-Ja L, Ahn J, Gronenborn AM, Polenova T (2012a) Solid-state NMR spectroscopy of protein complexes. In: Shekhtman A, Burz D (eds) Protein NMR techniques, Methods in molecular biology (Methods and Protocols). Humana Press, Totowa, NJ, p 831

Sun Y, Hays NM, Periasamy A, Davidson MW, Day RN (2012b) Monitoring protein interactions in living cells with fluorescence lifetime imaging microscopy. Methods Enzymol 504:371–391

Taipale M (2018) Two protein/protein interaction assays in one go. Mol Syst Biol 14(7):e8485. https://doi.org/10.15252/msb.20188485

Talavera D, Robertson DL, Lovell SC (2013) The role of protein interactions in mediating essentiality and synthetic lethality. PLoS One 8(4):e62866

Tavernier J, Eyckerman S, Lemmens I, Van der Heyden J, Vandekerckhove J, Van Ostade X (2002) MAPPIT: a cytokine receptor-based two-hybrid method in mammalian cells. Clin Exp Allergy 32:1397–1404

Thornton JM (2001) The Hans Neurath Award lecture of The Protein Society: proteins—a testament to physics, chemistry, and evolution. Protein Sci 10:3–11

Tomizaki K, Usui K, Mihara H (2010) Protein-protein interactions and selection: array-based techniques for screening disease-associated biomarkers in predictive/early diagnosis. FEBS J 277(9):1996–2005

Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91

Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O (2008) Architectures and functional coverage of protein-protein interfaces. J Mol Biol 381:785–802

Ulrichts P, Lemmens I, Lavens D, Beyaert R, Tavernier J (2009) MAPPIT (mammalian protein–protein interaction trap) analysis of early steps in toll-like receptor signalling. Methods Mol Biol 517:133–144

Uyttendaele I, Lemmens I, Verhee A, De Smet A-S, Vandekerckhove J, Lavens D, Tavernier J (2007) Mammalian protein-protein interaction trap (MAPPIT) analysis of STAT5, CIS, and SOCS2 interactions with the growth hormone receptor. Mol Endocrinol 21(11):2821–2831

Van Leene J, Eeckhout D, Cannoot B, De Winne N, Persiau G, Van De Slijke E et al (2015) An improved toolbox to unravel the plant cellular machinery by tandem affinity purification of Arabidopsis protein complexes. Nat Protoc 10:169–187

Vaynberg J, Qin J (2006) Weak protein–protein interactions as probed by NMR spectroscopy. Trends Biotechnol 24(1):22–27

Velazquez-Campoy A, Leavitt SA, Freire E (2004) Characterization of protein-protein interactions by isothermal titration calorimetry. In: Fu H (ed) Methods in molecular biology: protein protein interactions: methods and protocols, vol 261. Humana Press, Totowa, NJ, pp 35–54

Vizoso Pinto MG, Villegas JM, Peter J, Haase R, Haas J, Lotz AS, Muntau AC, Baiker A (2009) LuMPIS â A modified luminescence-based mammalian interactome mapping pull-down assay for the investigation of protein-protein interactions encoded by GC-low ORFs. Proteomics 9 (23):5303–5308

Völkel P, Le Faou P, Angrand P-O (2010) Interaction proteomics: characterization of protein complexes using tandem affinity purification–mass spectrometry. Biochem Soc Trans 38 (4):883–887

Vyncke L, Masschaele D, Tavernier J, Peelman F (2019) Straightforward protein-protein interaction interface mapping via random mutagenesis and mammalian protein protein interaction trap (MAPPIT). Int J Mol Sci 20(9):E2058. https://doi.org/10.3390/ijms20092058

Wang H, Liu C, Deng L (2018) Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting. Sci Rep 8(1):14285. https://doi.org/10.1038/s41598-018-32511-1

Wojcik J, Schächter V (2001) Protein-protein interaction map inference using interacting domain profile pairs. Bioinformatics 17(Suppl 1):S296–S305

Wu Y, Li Q, Chen X-Z (2007) Detecting protein–protein interactions by far western blotting. Nat Protoc 2(12):3278–3284

Xia B, Mamonov A, Leysen S, Allen KN, Strelkov SV, Paschalidis IC, Vajda S, Kozakov D (2015) Accounting for observed small angle X-ray scattering profile in the protein–protein docking server CLUSPRO. J Comput Chem 36:1568–1572

Xing S, Wallmeroth N, Berendzen KW, Grefen C (2016) Techniques for the analysis of protein-protein interactions *in vivo*. Plant Physiol 171:727–758

Xu X, Song Y, Li Y, Chang J, Zhang H, An L (2010) The tandem affinity purification method: an efficient system for protein complex purification and protein interaction identification. Protein Expr Purif 72:149–156

Xu Y, Piston DW, Johnson CH (1999) A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. Proc Natl Acad Sci U S A 96:151–156

Yamada M, Kabir MS, Tsunedomi R (2003) Divergent promoter organization may be a preferred structure for gene control in Escherichia coli. J Mol Microbiol Biotechnol 6(3–4):206–210

Ye P, Peyser BD, Pan X, Boeke JD, Spencer FA, Bader JS (2005) Gene function prediction from congruent synthetic lethal interactions in yeast. Mol Syst Biol 1(1):E1–E12

Zhang A (2009) Protein interaction networks: computational analysis. Cambridge University Press, New York, NY

Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T, Maniatis T, Califano A, Honig B (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature 490(7421):556–560

Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R et al (2001) Global analysis of protein activities using proteome chips. Science 293:2101–2105

Zuiderweg ERP (2002) Mapping protein−protein interactions in solution by NMR spectroscopy. Biochemistry 41(1):1–7

Thermodynamics of Protein-Ligand Binding

7

Komal S. Khatri, Priya Modi, Shilpa Sharma, and Shashank Deep

Abstract

Most of the biological processes consist of the interaction of the protein with a ligand. Understanding the molecular level picture of these interactions is vital for controlling these processes. In this context, the measurement of a complete set of thermodynamic parameters of interaction is a must. Here, we describe the importance of protein-ligand binding, the thermodynamic parameters describing the binding, and the methods to measure these parameters.

Keywords

Binding affinity · Entropy of binding · Enthalpy of binding · Heat capacity of binding · Isothermal calorimeter · NMR

7.1 Introduction

A variety of molecules such as ATP, GTP, saccharides, and macromolecules such as another protein, nucleic acids, and peptides are known to interact with protein and they are referred as a ligand. Some of the protein-ligand interactions are well known, such as the interaction between an enzyme with its substrate, the interaction of an antibody with its antigen, the interaction of protease with its inhibitor, interaction of ATPases with ATP, the interaction of GTPases with GTP, and interaction of transcription factors with DNA motifs.

Protein-protein interactions are essential for cellular regulation and signaling, apoptosis, cell division, cell proliferation, and cell division, to name a few. On the other side, the interaction of the molecules of the same protein results in aggregation

K. S. Khatri · P. Modi · S. Sharma · S. Deep (✉)
Department of Chemistry, Indian Institute of Technology Delhi, New Delhi, India
e-mail: sdeep@chemistry.iitd.ac.in

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_7

of the protein. This is the main cause of the rare diseases commonly known as amyloidosis (Cohen 1994a, b; Pepys et al. 1993; Sipe 1992). Parkinson, Alzheimer's, ALS, and Huntington are some of the examples of amyloid disease.

Protein-DNA interactions are crucial in controlling biological processes including regulation of gene expression, replication, chromosomal packing, recombination, and DNA repair, as well as, RNA transport and translation. Binding of proteins to DNA is important in genome maintenance and gene regulation. Gene expression is regulated by binding of the transcription factor, a protein, to specific DNA sequences. Protein-RNA interactions are important in the regulation of transcription and RNA processing.

The interaction of small molecules with protein is also important in biological processes and in the treatment of diseases. These molecules include both natural compounds in cells (metabolites) and chemically synthesized molecules (such as drugs). A group of proteins, known as enzymes, can catalyze various types of reaction of small molecules (substrates) through binding to them. Small molecules can also regulate protein functions through allosteric binding. Nucleotide compounds such as ATP, CTP, GTP, and TTP, deoxy nucleotide compounds such as dATP, dCTP, dGTP, and dTTP, cyclic nucleotides such as cAMP and cGMP, plant hormones, NAD group molecules such as NADH and NADPH, steroid hormones, and vitamins are all small-molecule metabolites with important regulatory roles. Examples of regulatory roles include regulation of oxygen binding to hemoglobin by 2,3-biphosphogylcerate, regulation of glycogen, sugar, and lipid metabolism by cAMP through binding to kinases, sterol regulation of protein kinases in yeast, and sphingolipid regulation of actin organizing protein. Understanding of interactions between proteins with peptides/drug molecules are needed to develop inhibitors of these interactions.

Several proteins possess intrinsic GTPase activity and have the ability to bind the guanine nucleotides, guanosine triphosphate (GTP) and guanosine diphosphate (GDP). Such proteins are basically G-proteins, which play a vital role in signal transduction as well as in various cellular processes such as cell growth, membrane vesicle transport, and protein synthesis. Similarly, several proteins possess intrinsic ATPase activity and have the ability to bind the adenosine triphosphate (ATP) and adenosine diphosphate (ADP). They catalyze the hydrolysis of phosphate bond of ATP and play an active role in energy conservation, active transport, and pH homeostasis.

In a nutshell, protein-ligand interaction is an integral part of every biological process. Therefore, it is important to understand the role of protein-ligand interactions to design a strategy for controlling various biological processes. On the basis of the type of ligand, some examples of protein-ligand interactions are illustrated in Table 7.1.

Table 7.1 Examples of protein-ligand interaction

S. no.	Protein-ligand	Examples
1.	Protein-protein	Transforming growth factors & their receptors (TGFβ-TβR2), HYHEL with lysozome, homodimer of sperm lysine, cytochrome c-cytochrome b5, RhoA and RhoGAP signaling complex, trypsin-Trypsin inhibitor, barnase and barstar complex, heterodimer of Bovine G protein
2.	Protein-peptide	Endothiapepsin-pepstatin, binding of S-protein with S-peptide, calmodulin-Melittin, Src SH2 domain-PYHmT phosphopeptide, binding of MDM2 to peptide p53 tumor suppressor transactivation domain
3.	Protein-DNA	DNA binding to histone protein in eukaryotic cell, DNA double helix with helicase DNA with DNA polymerase, DNA with endonuclease, binding of bent or distorted DNA with HMG proteins, interaction of CpG- rich viral DNA such as virus HSV-1, HSV-2, and murine cytomegalovirus with TLR9
4.	Protein-GTP/GDP	Guanylate binding protein-GTP, RAS-GTP, binding of GDP and GTP with G-protein
5.	Protein-ATP/ADP	ABC transporter-ATP, ATPase-ATP, protein kinase-ATP, HSP18-ATP, reticulocyte-binding protein-ATP

7.2 Biological Effects of Binding

After binding, a ligand may alter the protein functional behavior and thereby the biological processes in several different ways. Some of such effects are discussed below (Phizicky and Fields 1995).

1. Enzymes bind to their substrate to catalyze the reactions. For example, hexokinase binds to glucose and ATP to catalyze reaction between them. ATPases bind to ATP to hydrolyze them to ADP.
2. A large number of physiological processes are regulated through the differential binding of a ligand to different conformations of protein, which in turn regulates the structural transition among different conformations of the protein. Examples are known where a protein exists in both open and closed conformation, one of the conformations is stabilized on ligand binding. For example, differential binding of ATP to different conformations of cytochrome c affects the structural transition between them and thus regulates the apoptosis (Ahluwalia et al. 2011).
3. A protein may get activated or deactivated or destroyed on interaction with another protein. For example, a phage P22 repressor becomes inactive upon interaction with its antirepressor (Vershon et al. 1987). The interaction of trypsin with trypsin inhibitor makes it inactive (Savage and Morrison 2003). On the other hand, small GTPase "*RAS*" becomes active when bound to GTP; however after activation, ATP is released and binding takes place with ADP (Sprang 2016). A glycoprotein "*Fibronectin*" binds with transforming growth factor (TGF-β) and enhances its bioactivity (Dallas et al. 2005). Antibodies interact with viruses or

bacteria to mark them for destruction. Macrophage Inflammatory Proteins (MIP) get activated on binding with cytokines like IFN-gamma or bacterial endotoxins like LPS and undergo structural changes which help to kill the invaded bacteria and infected cells (Arango Duque and Descoteaux 2014).

4. Protein-ligand interactions may facilitate the further binding of another ligand by creating a new binding site. This is known as cooperative binding. For example, binding of transforming growth factor beta (TGF-β) with its receptor type II creates a binding site for its binding with receptor type I (Hart et al. 2002). Examples of several multi-subunit proteins exist where binding site exists at the interface of two subunits. ADP binding site exists at the interface of α- and β-subunit of F_1-ATPase from *E. coli*. Also, hemoglobin binds with O_2 facilitates further binding of O_2 molecules.
5. The protein-ligand interactions may alter the specificity of a protein for its substrate. For example, the binding of lactalbumin with lactose synthase decreases its K_m for glucose by 1000-fold (Hill and Brew 1975). Also, the interaction of succinate thiokinase and α-ketoglutarate dehydrogenase reduces the K_m for succinyl coenzyme A by 30-fold (Porpaczy et al. 1983).
6. The ligand-receptor interaction is very important for the signal transduction process to start. For example, only after the binding of transforming growth factors beta with its receptor type II and receptor type I, the intracellular part undergoes phosphorylation leading to signal transmission.
7. Binding of a ligand regulates the subunit interaction within a protein which may generate positive, negative, and mixed cooperativity. 2,3-biphosphogylcerate regulates the oxygen binding to hemoglobin.

7.3 Specificity and Selectivity in Protein-Ligand Interaction

Specificity and selectivity are particularly important in signal transduction since signals from two homologous receptors should not interfere with each other. A given cell can express more than 100 different such receptors with high sequence and structural homology. Thus, the interaction of a ligand to a receptor should be much stronger than other homologous receptors.

7.4 Thermodynamics of Protein-Ligand Interaction

The protein interacts with its ligand with high affinity and the interaction is specific. Apart from structural information, it is important to understand the thermodynamics of the protein-ligand interaction to get a complete molecular picture of affinity and specificity. Measurement of thermodynamic parameters is important in the understanding of specificity and selectivity in protein interaction (Du et al. 2016). For specific interactions, the affinity of a specific ligand to a receptor should be approximately three orders of magnitude stronger than a nonspecific interaction (Ladbury 2010).

7.4.1 Binding Affinity

It is a parameter measured to understand the strength of protein-ligand interaction. The association of protein (P) and ligand (L) is given as Eq. (7.1):

$$P + L \rightleftharpoons PL \tag{7.1}$$

If the rate constant for the forward step, i.e., bimolecular association, is given by k_a and for the reverse step, i.e., dissociation, the rate constant is given by k_d, then the rate of complex formation is given as:

$$\frac{d[PL]}{dt} = k_a[P][L] - k_d[PL] \tag{7.2}$$

where $[PL]$, $[P]$, and $[L]$ are the concentration (in mol dm^{-3}) of complex PL and the free protein P and ligand L, respectively. At thermodynamic equilibrium, the rate of complex formation is zero and hence

$$\frac{[P]_{eq}[L]_{eq}}{[PL]_{eq}} = \frac{k_d}{k_a} = \frac{1}{K_a} = K_d \tag{7.3}$$

where K_A represents the affinity constant while K_d is referred as the dissociation constant. K_d is expressed in mM, μM, or nM. Lower the value of K_d, higher will be the affinity. Therefore, if dissociation constant is in the nanomolar (nM) range, the ligand will bind to a particular protein more tightly as compared to another ligand for which dissociation constant of complex is in μM range. Some of the examples of the different strengths of the protein-ligand interactions are given in Table 7.2.

7.4.2 Free Energy Change

Protein-ligand interactions are accompanied with the changes in the enthalpy (ΔH), entropy (ΔS), free energy (ΔG), and heat capacity (ΔC_p) of the system. Hill (1986)

Table 7.2 Examples of protein-ligand interactions of different strength

S. no.	Dissociation constant (K_d)	Strength of interaction	Examples
1.	<0.1 nM	Very strong	Antibody-antigen interaction, PDBu binding with protein kinase C
2.	100–0.1 nM	Strong	TβR2-TGFβ3 interaction, Riboflavin binding protein-vitamin B2 interaction, human cyclophilin 18 (hCyp18) with cyclosporine A (CsA) interaction
3.	10 μM–100 nM	Medium	Enzyme-substrate reaction, TtgR protein binding with naringenin and phloretin
4.	>10 μM	Weak	Cytochrome c- cytochrome b5, Rho-BCM interaction, bacterial MreB protein-Sceptrin interaction

and Sommerfeld (1964) discussed the thermodynamic basis of the Eq. (7.3). For standard conditions, i.e., at 1 atm pressure, the Eq. (7.3) can be modified as:

$$\frac{[P]^0[L]^0}{[PL]^0} = \frac{K_d}{c_0} = \exp\left(-\Delta G^0/RT\right) \tag{7.4}$$

The association/dissociation constants (K) is related to the *standard free energy change* for the protein-ligand interaction (ΔG^0).

$$\Delta G^0 = -RT \ln\left(\frac{K_d}{c_0}\right) \tag{7.5}$$

where R is gas constant, T is the temperature, and c_0 is the concentration chosen to define the standard state and usually equal to 1 M.

7.4.3 Enthalpy Change

Protein-ligand interaction involves the formation of new bonds between protein and ligand as well as breaking of existing bonds between protein/ligand and solvent which includes formation or breaking of ionic bonds, hydrogen bonds, electrostatic interactions. Heat is needed to break a bond while heat is released when a bond is formed. This intake or release of heat energy on bond breaking/forming contributes to the enthalpy change for a protein-ligand formation/dissociation process. If the net intake of heat is more than net release of heat in binding process, the process is endothermic. On the other hand, if the net intake of heat is less than the net release of heat in the binding process, the process is exothermic. Enthalpy change is a static component of the free energy change of the interaction. Mathematically, the free energy change is expressed as:

$$\Delta G = \Delta H - T\Delta S \tag{7.6}$$

7.4.4 Entropy Change

The second factor affecting the free energy change of binding is the entropy change. It is a dynamic contribution towards free energy change. There are two different types of entropy: conformational entropy and solvation entropy.

7.4.4.1 Conformational Entropy

Binding of ligand to a protein involves a number of conformational changes. This also brings many rotational and translational changes in protein resulting in the stabilization of a particular conformation at the interface. In this process, the entropy of the system decreases. However, in cases, the entropy of the ligand binding has

shown to get increased due to the increase in number of conformational degrees of freedom. By using NMR and MD simulation, MacRaild et al. showed that there is entropic cost of binding of d-galactose to arabinose binding protein (MacRaild et al. 2007). On the other hand, Stockmann et al. showed that dynamics increases on the binding of *p*-(gly)*n*-substituted benzenesulfonamide to bovine carbonic anhydrase II (Stockmann et al. 2008). Tzeng and Kalodimos showed that the binding process is regulated with the changes in the fast and slow internal dynamics (Tzeng and Kalodimos 2012).

7.4.4.2 Solvation Entropy

This involves the displacement of water molecules from the binding interface. This is also called the hydrophobic effect and results in an increase in entropy. Therefore, in such cases, the binding is favored by an increase in entropy.

Enthalpy-entropy compensation: Protein ligand interaction is usually associated with enthalpy entropy compensation. If the binding leads to more and/or tighter salt bridges, hydrogen bonds, and van der Waals interaction, this results in a loss in the conformational flexibility of the residues at the interface. Thus, a negative enthalpy change is compensated by a decrease in conformational entropy. Freire and coworkers found that introducing a hydrogen bond acceptor into an HIV-1 protease inhibitor resulted in a 3.9 kcal mol^{-1}gain in the enthalpic contribution to binding but was completely offset by a corresponding loss in the entropic contribution, resulting in no net change in affinity (Lafont et al. 2007). Chodera and Mobley (2013) reviewed the various physical origin of compensation proposed. Some of the most prevalent ones are: (a) solvent reorganization on binding and (b) conformational restriction of bound protein and ligand.

7.4.5 Heat Capacity Change

The heat capacity change (ΔC_p) is another thermodynamic quantity, which gets affected during protein-ligand binding. It is obtained by measuring the enthalpy change at different temperatures. The heat capacity change for a dissociation process is given as:

$$\Delta C_p = \frac{d(\Delta H_d)}{dT} = T\frac{d(\Delta S_d)}{dT} \tag{7.7}$$

7.5 Type of Interactions and Their Effect on Enthalpy and Entropy of the Binding

A protein binds with a ligand through a number of non-covalent interactions. The most common interactions and their effect on the thermodynamic parameters of binding are as follows.

7.5.1 Hydrophobic Interactions

Hydrophobic interactions can be described as the tendency of nonpolar substances to come together to exclude water molecules and form an aggregate in an aqueous solution. The idea of hydrophobic interaction was put forward by Kauzmann in the context of protein stability, diverting the attention from the role of hydrogen bonding which was considered to be the most important force contributing to protein stability at that time. Protein folding and protein-ligand interaction are thermodynamically driven process. Transfer of a polar group into water is accompanied by negative enthalpy and entropy change. Transfer of apolar group into water is also accompanied by negative enthalpy and entropy change. Decrease in the entropy of water is ascribed to an increase in ordering of water around the apolar group. Thus, the apolar group tends to aggregate in water since it leads to the expulsion of water.

Protein-protein interfaces are commonly hydrophobic and composed of a large number of nonpolar residues. When a protein folds or interacts with ligand, its nonpolar side chains come together to exclude water in order to minimize the ordering of water near their surface. These hydrophobic interactions result in a gain in solvation entropy and hence decrease in free energy and thereby stabilizing the protein-protein complexes.

The sign of the heat capacity change provides an idea about the hydrophobicity of the core of the protein-ligand complex. For example, the burial of hydrophobic surfaces during protein-protein interaction is indicated by a large negative value of the heat capacity change (Fersht 1972); although if the polar surfaces are buried in nonpolar environment, then the heat capacity change for the interaction is positive but of lower magnitude (Walker 1978). Better binding of FK506 to the FK506 binding protein (FKBP-12) relative to the Y82F variant is driven by the favorable entropy associated with the release of water molecules on binding (Pearlman and Connelly 1995).

7.5.2 Electrostatics Interactions

Electrostatic interactions are formed by the ionizable residue side chains (His, Asp, Glu, Lys, and Arg), or the free ionized groups of the amino- and carboxy-termini. The electrostatic interaction energy between two charged atoms is given by Coulomb's law.

$$E_{\mathrm{el}}\left(\text{in kcal mol}^{-1}\right) = 332.06\left(\frac{q_1 q_2}{\varepsilon r_{12}}\right) \tag{7.8}$$

where q_1 and q_2 are charges on the atoms involved in electrostatic interaction, r_{12} is the distance between them, and ε is dielectric constant of the medium. The electrostatic interaction energy between two atoms at a distance of 3 Å with single positive and negative charge, respectively, is 1.4 kcal mol^{-1}.

Electrostatic interactions are even effective at relatively long ranges, particularly in a low dielectric constant environment. The strength of such interaction is inversely related to the dielectric constant of the environment resulting in stronger electrostatic interaction in nonpolar solvents than in aqueous solvents. The dielectric constant of nonpolar protein interior is around 2–4.

Electrostatics has both stabilizing and destabilizing effects on protein-protein interaction. Protein interaction with ligand will lead to the formation of hydrogen bond, ion pair between them leading to the release of heat. There is the breaking of protein-water and ligand-water bonds on interaction, leading to the absorption of heat. The interaction will also lead to the burial of charges, ion pairs, and hydrogen bonds and thus destabilize the complex formation. The electrostatic-driven process is generally dependent on pH and/or ionic strength. The total charge on the protein depends on pH and the extent of interaction among these charges gets affected in the presence of salts since salts shield charges. An estimated value of 1–3 kcal $mol^{-1}$for the stabilization per ion pair results in a value of 5–15 kcal mol^{-1}stabilization.

7.5.3 van der Waals Interaction (Dispersion Force)

van der Waals interactions are short range and weak forces resulting from interactions among fixed or induced dipoles. These interactions may lead to attraction or repulsion between two atoms depending on the distance between them. If atoms are extremely close, the repulsion will take place. These interactions are generally modeled by Lennard-Jones potential.

$$E_{i,j} = \epsilon \left[\left(\frac{r_{\min}}{r} \right)^{12} - 2 \left(\frac{r_{\min}}{r} \right)^{6} \right] \tag{7.9}$$

$E_{i,\,j}$ is the van der Waals interaction energy between atoms i and j at a distance r apart. $r_{\min}$ is equilibrium distance between the two atoms, ϵ is the depth of potential well. It involves the interaction of the nonpolar side chains in the interior of the protein having minimal contact with water. van der Waals energy of a pair of atoms is about 0.1–0.2 kcal mol^{-1}. Such interactions are numerous and additive. The packing density of the protein cores is determined by the strength of these interactions.

7.5.4 Hydrogen Bonding

A hydrogen bond is formed when a hydrogen atom covalently bound to an electronegative atom (donor) is in close proximity to an electronegative atom (acceptor).

$$E_{\mathrm{HB}} = \epsilon_{\mathrm{HB}} \left[5 \left(\frac{r_{\min}}{r} \right)^{12} - 6 \left(\frac{r_{\min}}{r} \right)^{10} \cos^{4}\theta \right] \tag{7.10}$$

E_{HB} is energy of formation of hydrogen bond between donor and acceptor separated by distance r. r_{min} is equilibrium distance between donor and acceptor,ϵ_{HB} is the depth of potential well, and θ is the donor-hydrogen-acceptor angle. The protein-ligand binding will involve both breaking of hydrogen bond between solvent and protein/ligand and forming of a hydrogen bond between protein and ligand. Hydrogen bond formation is accompanied by heat release whereas breaking of the bond is accompanied by heat absorption.

7.5.5 Enthalpy/Entropy-Driven Protein-Ligand Interaction

The process of protein-ligand binding can be either enthalpy driven or entropy driven based on the phenomenon of enthalpy-entropy compensation (Cooper et al. 2001). As shown in Fig. 7.1, case (a) illustrates the enthalpy-driven process where the binding is favored by the formation of hydrogen bonds and other non-covalent bonds. On the other hand, negative entropy indicates the loss of conformational freedom on the binding. This is the most observed case in protein-ligand binding. The case (b) depicts the entropy-driven process with positive entropy which indicates hydrophobic interaction taking a more predominant role in binding whereas positive enthalpy indicates loss of hydrogen bonds or non-covalent interaction during the binding process. The last part of the figure, case (c), indicates that the process of binding is favored by both enthalpy and entropy.

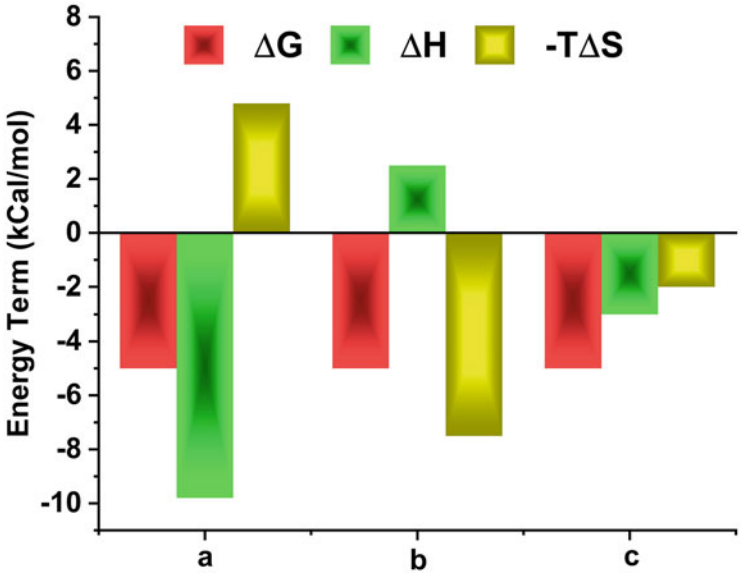


Fig. 7.1 Representation of thermodynamic data observed via ITC

Frasca showed the binding of three recombinant antibodies with the same antigen having similar K_d values but different enthalpy change and entropy change profile (Frasca 2016). In the interaction between bisphosphonate and enzyme farnesyl diphosphate synthase, the side chain of the biphosphonate decides whether its binding with the enzyme will be enthalpy or entropy driven. If the side chain is negatively charged, then the binding is entropy driven; on the other hand, if it has a neutral side chain, then binding will be enthalpy driven (Yin et al. 2006). Similarly, the binding of opioid peptides to human dipeptidyl peptidase III is entropy-driven process (Bezerra et al. 2012).

7.6 Measurement of Binding Affinity

7.6.1 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a very powerful tool to study the thermodynamic aspects of protein-ligand interactions existing in numerous biological systems (Cooper 1999; Leavitt and Freire 2001). ITC measurement helps us to determine actual enthalpy change in binding in contrast to van't Hoff enthalpy obtained with spectroscopy. It also gives the binding affinity (K_b or dissociation constant, K_d), ΔG^0, ΔS^0, and number of binding sites available on the protein (n) (Holdgate 2001). Modern ITC technologies can measure the dissociation constant (K_d) for the complexes with high affinity binding ($K_d \sim$ nM range) as well as weak affinity binding ($K_d \sim$ in mM range) (Frasca 2016).

7.6.1.1 Protein with One Set of Binding Site

In ITC technique, for a simple protein-ligand binding process:

$$P + L \rightleftharpoons PL,$$

$$[PL] = K_b[P][L] \tag{7.11}$$

$[P]$, $[L]$, and $[PL]$ are the concentration of free protein, free ligand, and bound protein, respectively, and K_b is the association constant.

If we titrate protein with ligand, a part of protein will get complexed. The total protein concentration, $[P]_T$, is sum of free $[P]$ and bound protein $[PL]$

$$[P]_T = [P] + [PL] \Rightarrow [P] = [P]_T - [PL] \tag{7.12}$$

Similarly, total ligand concentration, $[L]_T$, is expressed as:

$$[L]_T = [L] + [PL] \Rightarrow [L] = [L]_T - [PL] \tag{7.13}$$

Therefore, concentration of bound protein is given as:

$$[PL] = K_b([P]_T - [PL])([L]_T - [PL]) \tag{7.14}$$

Rearranging the equation form a quadratic equation in terms of $[PL]$ as:

$$[PL]^2 - \left(\frac{1}{K_b} + [P]_T + [L]_T\right)[PL] + [P]_T[L]_T = 0 \tag{7.15}$$

Dividing Eq. (7.15) with $[P]_T{}^2$, we get

$$\frac{[PL]^2}{[P]_T{}^2} - \left(\frac{1}{K_b[P]_T} + 1 + \frac{[L]_T}{[P]_T}\right)\frac{[PL]}{[P]_T} + \frac{[L]_T}{[P]_T} = 0 \tag{7.16}$$

The fraction of sites occupied by the ligand L is given as

$$f = \frac{[PL]}{[P]_T}$$

Using this relation in Eq. (7.16), we get

$$f^2 - \left(1 + \frac{1}{K_b[P]_T} + \frac{[L]_T}{[P]_T}\right)f + \frac{[L]_T}{[P]_T} = 0 \tag{7.17}$$

This is a quadratic equation in terms of f and the solution is given as:

$$f = \frac{\left(1 + \frac{1}{K_b[P]_T} + \frac{[L]_T}{[P]_T}\right) - \sqrt{\left(1 + \frac{1}{K_b[P]_T} + \frac{[L]_T}{[P]_T}\right)^2 - \frac{4[L]_T}{[P]_T}}}{2} \tag{7.18}$$

Now assuming,

$L_R = \frac{[L]_T}{[P]_T}$ = absolute ratio of concentrations of ligand to protein at any point of the titration and

$$r = \frac{1}{[P]_T K_b} = \frac{K_d}{[P]_T}$$

Therefore, the modified solution of the quadratic equation will be

$$f = \frac{(1 + r + L_R) - \sqrt{(1 + r + L_R)^2 - 4L_R}}{2} \tag{7.19}$$

Total heat released in the system is given as:

$$Q = [PL] \cdot \Delta H^0 \cdot V_0 \tag{7.20}$$

or

$$Q = f.[P]_T \cdot \Delta H^0 \cdot V_0 \tag{7.21}$$

where ΔH^0 = standard enthalpy change for the interaction

V_0 = working volume of the sample cell

Substituting the solution of f in Eq. (7.21), we get

$$Q = \frac{[P]_T \cdot \Delta H^0 \cdot V_0}{2}\left[(1 + r + L_R) - \sqrt{(1 + r + L_R)^2 - 4L_R}\right] \tag{7.22}$$

The important factor in an ITC curve is the constant c and given as:

$$c = \frac{1}{r} = K_b[P]_T = \frac{[P]_T}{K_d} \tag{7.23}$$

If there are n binding sites on the protein, then

$$L_R = \frac{[L]_T}{n[P]_T} \text{ and}$$

$$r = \frac{1}{n[P]_T K_b} = \frac{K_d}{n[P]_T}$$

$$c = nK_b[P]_T = \frac{n[P]_T}{K_d} \tag{7.24}$$

The plot obtained in ITC calculation is sigmoidal. The most important factor affecting the sigmoidal shape of the resulting curve is c. It depicts the nature of the binding affinity of the respective protein-ligand interaction. Wiseman et al. (1989) showed that the value of c must be in a range of 10–500 for the determination of accurate value for thermodynamic parameters for the respective protein-ligand interaction. Below $c < 10$, the resulting curve becomes more and more linear and resulting values of thermodynamic parameters are not feasible.

Later on, Turnbull and Daranas (2003) did some modifications in the equation. They modified the calculation of concentration of protein and ligand at any time t as:

$$[P]_T = [P]_0\left(\frac{1 - \frac{\Delta V}{2V_0}}{1 + \frac{\Delta V_i}{2V_0}}\right) \text{ and } [L]_T = \frac{\Delta V_i[L]_{syr}}{V_0}\left(1 - \frac{\Delta V_i}{2V_0}\right) \tag{7.25}$$

where $[P]_0$ = initial protein concentration, ΔV_i = total volume of ligand added on ith injection, $[L]_{syr}$ = concentration of ligand in the syringe

They proved that for a simple 1:1 binding model, the values of affinity constants and standard free energy change can be determined accurately. They calculated the binding constant of the complex of Con A with methyl α-mannopyranoside at 27 °C to be 6890 M^{-1} with a confidence interval of ±320 M^{-1} even with the small value of $c = 0.04$.

Tellinghuisen showed that ITC can also be used even at a very low value of c (Tellinghuisen 2008). Earlier it was shown that the main cause of inaccurate results is the poorly known n value. Tellinghuisen proved that the accuracy of binding constants (K_a) is independent of the n values at low c. Later, Biswas and Tsodikov showed that the most optimal c value for determination of the reliable K_d and ΔH lie in between 5 and 20 (Biswas and Tsodikov 2010). This was further revised by Broecker et al. (2011). He did Monte Carlo simulations for precision in the measurement of ΔH and n, and deduced the c should be greater than 40.

7.6.1.2 Two or More Sets of Independent Binding Sites

Many cases involve two or more independent binding sites available on protein for the ligand. Le et al. (2013) developed an algorithm to account for such interactions. For such protein-ligand complexes,

$$P + nL \rightleftharpoons PL_n \tag{7.26}$$

the total ligand concentration is expressed as:

$$L_T = [L] + (PL + 2PL_2 + 3PL_3 + \ldots + nPL_n)$$

$$L_T = [L] + P_T\left(1\,\frac{PL}{P_T} + 2\frac{PL_2}{P_T} + 3\frac{PL_3}{P_T} + \ldots + n\frac{PL_n}{P_T}\right)$$

$$L_T = [L] + P_T\sum_{i=1}^{n}(n_i m_i) \tag{7.27}$$

where $[L]$ = free ligand concentration
n = number of binding sites
m = fraction of occupied binding sites
and the binding constant is given as:

$$K_i = \frac{m_i}{(1 - m_i)[L]} \tag{7.28}$$

or

$$m_i = \frac{[L]K_i}{1 + [L]K_i} \tag{7.29}$$

Solving these equations for free ligand concentration leads to $(n + 1)^{th}$ degree polynomial equation. After solving for $[L]$, the total heat content (Q_i) after ith injection is expressed as:

$$Q_i = P_T V_o \sum_{i=1}^{n} n_i m_i \Delta H_i \tag{7.30}$$

m_i can be calculated as in Sect. 7.6.1.1.

7.6.1.3 Example of Study of Thermodynamics of Protein-Ligand Interaction Using ITC

Nall et al. illustrated the thermodynamics of the interaction of horse heart cytochrome (cyt) c with two different monoclonal antibodies (MAb) 2B5 and 5F8 (Pierce et al. 1999). Horse heart cyt c is an example of protein with independent noninteracting binding sites. MAb 2B5 covalently binds to the site having critical residue Pro 44, while MAb 5F8 binds to the opposite site where lysine residue is present at position 60. It was reported that MAb 5F8 ($K_d = 0.5$ nM) binds with cyt c more strongly than MAb 2B5 ($K_d = 0.07$ nM). The free energy and enthalpies of the interaction of cyt c with 2B5 ($\Delta G_b^0 = -12.6$ kcal mol^{-1}) and 5F8 ($\Delta G_b^0 = -13.9$ kcal mol^{-1}) are similar and favorable. However, there is decrease in entropy found in binding with MAb 2B5 ($\Delta S_b^0 = -28.2$ cal mol^{-1} K^{-1}) and MAb 5F8 ($\Delta S_b^0 = -26.3$ cal mol^{-1} K^{-1}) due to conformational restrictions of side chain in the formed complex. At 25 °C for MAb 5F8, the free energy of binding showed negligible dependence on temperature due to thermal independence of the enthalpic (ΔH_b^0) and entropic contribution ($-T\Delta S_b^0$).

On the other hand, for MAb 2B5 binding, the free energy of binding is also independent of temperature but due to compensating effect of enthalpic (negative change with temperature) and entropic terms (positive change with temperature). The probable difference between the two binding constants is depicted by the difference between the heat capacity change (ΔC_p^0) for the two complex formations. As we discussed in earlier sections, heat capacity change is estimated by the thermal dependence of the enthalpy change for the process. Here, as mentioned for 2B5-cyt c complex, the enthalpy change shows a strong negative thermal dependence, and therefore the calculated heat capacity change was determined to be -580 cal mol^{-1} K^{-1}. While for MAb 5F8-cyt c complex, heat capacity change was calculated as -172 kcal mol^{-1} K^{-1}.

In brief, ITC provides the structural information of the protein-ligand complexes and also describes thoroughly the interactions of these complexes.

7.6.2 Methods Based on Measurement of Fraction of Bound Protein

The binding affinity of a protein-ligand complex is measured by measuring fraction of bound protein or ligand. Here, first the protein is titrated with its ligand and then the fraction of the bound protein is measured. Fraction of bound protein is given by

$$\text{Binding Affinity } (\nu) = \frac{[\text{Bound protein}]}{[\text{Total protein}]} = \frac{[\text{Bound protein}]}{[\text{Free protein}] + [\text{Bound protein}]}$$

$$\nu = \frac{[PL]}{[P] + [PL]}$$

7.6.2.1 Experimental Tool to Measure Fraction of Protein Bound to Ligand

The experimental techniques used to measure the fraction of bound protein in the protein-ligand complex are either based on the separation of the complex or on the detection of a complex. These are given as below.

Techniques Based on Separation of Complex

The most appropriate technique used for this study is gel filtration chromatography. In this technique, the molecules are separated according to their size. The molecule with higher molecular weight elutes first and the one with lower molecular weight protein coming later. This technique is very useful if the size of protein and complex differs considerably resulting into well separated peaks of protein, ligand, and complex.

Techniques Based on Detection of Complexes

The techniques used to monitor the complex formation are used to determine the bound fraction of the protein. Some of these techniques are illustrated below.

Equilibrium Dialysis

This technique is used to measure free and bound ligand. Here, a dialysis membrane is used through which free ligand can diffuse but bound ligand cannot. When equilibrium is attained, the concentration of free ligand on either side of the membrane becomes equal and therefore can be measured accurately by spectroscopic techniques. Once the concentration of a free ligand is known, the amount of bound protein can be calculated easily.

Fluorescence Methods (Quenching, Anisotropy, FRET)

The study of protein-protein interactions can be achieved using numerous fluorescence-based approaches. The basic requirement for this study is the presence of fluorophore in either ligand or protein or both. Binding of a ligand is generally used whose photophysical parameters change in free and bound form. The different photophysical parameters used for estimation of the protein interactions are polarization, lifetime, average energy, and quantum yield. Fluorescence of the protein-ligand complex can be measured using steady-state or time-resolved techniques (Yan and Marriott 2003a, b).

Steady fluorescence techniques, based on intrinsic and extrinsic fluorescence intensity, are quite routinely used to get the dissociation constant and thermodynamic parameters of binding. Fluorescence intensity change of protein on the

addition of ligand can be used to calculate the fraction of bound protein which is related to dissociation constant. The study of protein-ligand binding can be done by the analysis of time-resolved emission decay data from the excited state of a fluorescent probe. The decay rate of free and bound protein can be used to resolve both the species. The measurement of lifetime of decay in the presence and absence of the ligand can be used to calculate the fraction of bound protein and thus dissociation constant of the complex. FLIM is a microscope which records and obtains the images in terms of decay-time or lifetime. Based on the lifetime images of the donor probe, the interaction between the proteins present in cells can be easily mapped (Clegg et al. 2003; Harpur et al. 2001).

Other techniques used to determine the hydrodynamic properties of the protein-protein interactions are fluorescence polarization (FP) or fluorescence anisotropy (FA) (Clayton et al. 2002; Fowler et al. 2002; Jameson et al. 2003; Subramaniam et al. 2003; Turconi et al. 2001). Various protein-protein interactions within living cells can be studied using microscope-based imaging of fluorescence anisotropy (Clayton et al. 2002; Gautier et al. 2001; Ventre et al. 2003; Yan and Marriott 2003a, b). Another technique used to study protein-protein interactions is fluorescence correlation spectroscopy (FCS). This technique is very efficient while working with low protein concentration (nanomolar scale) and low sample volume (femtoliter scale). It is based on the analysis of intensity fluctuation of labeled proteins (Yan and Marriott 2003a, b).

Fluorescence resonance energy transfer (FRET) is a physical phenomenon, which involves the non-radiative transfer from the excited state of a fluorophore of a donor protein to another fluorophore attached to a binding partner (acceptor). This technique helps to measure the distance between the donor and the acceptor probes and thereby the proximity between two proteins. Due to the FRET phenomenon, the quantum yield and the lifetime of the donor protein decrease upon the formation of protein complexes.

Nuclear Magnetic Resonance Technique

In the nuclear magnetic resonance (NMR) technique, the probe can be a ligand or protein or both. For a protein–ligand system with high binding affinity and slow exchange on the chemical shift time scale, resolved signals might be expected for the free and bound states. In these cases, integration of the distinct resolved signals will give the resulting values for $[L]/[P]$ and $[PL]$. On the other hand, for a system where exchange on the chemical shift scale is fast, the observed NMR response of a ligand/protein is the mole fraction weighted average of the NMR parameters of the free and bound states. In these cases, the dissociation constant (K_d) can be calculated as discussed in the next section (Fielding 2003).

7.6.2.2 Calculation of Fraction of Bound Protein from Signal Obtained from Spectroscopic Technique

In a spectroscopic technique, the observed signal is the contribution from protein, ligand, and the complex.

$$\text{Total signal } (S) = \text{signal from protein} + \text{signal from ligand} + \text{signal from complex}$$

$$S_T = S_p[P] + S_L[L] + S_{PL}[PL] \tag{7.31}$$

where S_P = signal from per molar free protein,

S_L = signal from per molar free ligand,

S_{PL} = signal from per molar bound protein

Case (1): When the signal (S) at a given wavelength is from bound protein or bound ligand only; for example, binding of BSA (P) with ANS (L). ANS is a fluorescent probe which binds only with the hydrophobic core of the protein and shows emission maxima around 475 nm, while free ANS in the buffer does not show any signal. Proteins, including BSA, generally emit at 370 nm due to aromatic amino acid, tryptophan. Thus, at 475 nm, the fluorescence signal from free BSA (P) and free ANS (L) is zero.

$$S_P = S_L = 0 \tag{7.32}$$

Therefore, the total signal obtained at 475 nm will be only from the bound protein, i.e., BSA-ANS complex.

$$S_T = S_{PL}[PL] \tag{7.33}$$

At infinite time, the signal (S_∞) corresponds to the maximum concentration of ANS-bound BSA which is equal to total protein concentration. The signal is represented as

$$S_\infty = S_{PL}[PL]_{\max} = S_{PL}P_T \tag{7.34}$$

On rearranging, we get

$$\frac{[PL]}{P_T} = \frac{S_T}{S_\infty} \tag{7.35}$$

Case (2): When the signal is from free and bound protein only: This is a case when a ligand is not a fluorophore or does not have any signal in the specified wavelength region. For example, if the fluorescence study of BSA and ANS binding is performed at 370 nm.

At the start of the titration, the signal corresponds to total free BSA only.

$$S_0 = S_P P_T \tag{7.36}$$

On addition of ANS to BSA sample, the total signal will have contribution from both protein (BSA) and ligand-bound protein (BSA-ANS).

$$S = S_P[P] + S_{PL}[PL] \tag{7.37}$$

The concentration of free BSA is the difference between the total protein concentration and protein in bound state.

$$S = S_P(P_T - [PL]) + S_{PL}[PL] \tag{7.38}$$

As in the previous case, at high enough ligand concentration, the signal becomes constant and total signal at that point corresponds to a maximum concentration of bound complex or total protein concentration as

$$S_\infty = S_{PL}[PL]_{\max} = S_{PL}P_T \tag{7.39}$$

The subtraction of S and S_0 gives

$$S - S_0 = [PL](S_{PL} - S_P) \tag{7.40}$$

Similarly,

$$S_\infty - S_0 = P_T\left(S_{PL} - S_p\right) \tag{7.41}$$

On arranging these relations, we get

$$\frac{[PL]}{P_T} = \frac{S - S_0}{S_\infty - S_0} \tag{7.42}$$

Case (3): When the signal is from the ligand and bound ligand only: When the signal has a contribution from the only ligand either in the free or bound state, then at any time of the process, the concentration of ligand-bound protein can be determined as

$$\frac{[PL]}{L_T} = \frac{S - S_0}{S_\infty - S_0} \tag{7.43}$$

For example, curcumin (L) shows a weak fluorescence signal at 522 nm (excitation wavelength at 420 nm) in phosphate buffer which enhances upon binding with protein, BSA (A). However, in this wavelength range BSA (P) does not show any fluorescence (Nadi et al. 2015).

7.6.2.3 Relationship Between Fraction of Protein Bound to Ligand and K_d

Protein with One Set of Binding Site

The binding or dissociation constant (K_a or K_d) of protein-ligand complex can be related to the fraction of the bound protein as:

$$f = \frac{K_a[P][L]}{[P] + K_a[P][L]} = \frac{K_a[L]}{1 + K_a[L]} \tag{7.44}$$

or

$$f = \frac{[L]}{K_d + [L]} \tag{7.45}$$

When the fraction of the bound protein is 0.5, then Eq. (7.45) can be modified as

$$2[L] = K_d + [L] \tag{7.46a}$$

or

$$[L] = K_d \tag{7.46b}$$

Therefore, the dissociation constant (K_d) is defined as the concentration of the ligand which saturates half of the total sites on protein.

Assumption: if the total protein concentration is assumed to be very low, i.e., $[P]_{total} << K_d$, then, the concentration of the ligand will be

$$[L] \cong [L] + [PL] \cong [L]_{total} \tag{7.47}$$

Therefore, using Eqs. (7.45) and (7.47) the fraction of bound protein is calculated as

$$f = \frac{[L]_{Total}}{K_d + [L]_{Total}} \tag{7.48}$$

Figure 7.2 shows the change in the fraction of bound protein with ligand concentration. The change in bound protein concentration changes exponentially with an increase in ligand concentration.

On rearranging the Eq. (7.45), we get

$$f[L] + fK_d = [L] \tag{7.49}$$

Dividing Eq. (7.12) by $[L]K_d$, we get

$$\frac{f}{K_d} + \frac{f}{[L]} = \frac{1}{K_d} \tag{7.50a}$$

or

$$\frac{f}{[L]} = \frac{1}{K_d} - \frac{f}{K_d} \tag{7.50b}$$

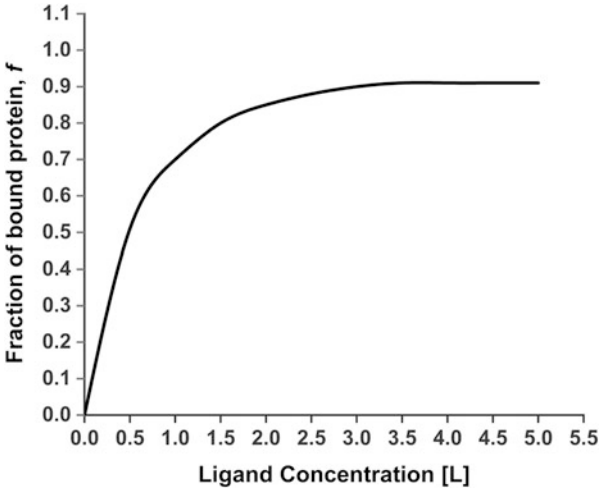


Fig. 7.2 Plot of fraction of protein bound versus concentration of ligand

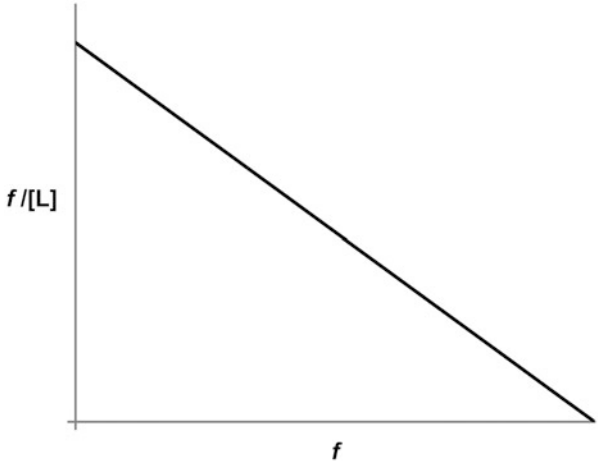


Fig. 7.3 Scatchard plot: plot of $\frac{f}{[L]}$ versus f

When $f/[L]$ is plotted against f we get a straight line. The slope of the line is equal to $-(1/K_d)$ and intercept is $(1/K_d)$. This plot is known as *Scatchard plot* and shown in Fig. 7.3.

Protein Having more than One Binding Site

Protein-ligand binding is also affected by the number and nature of the binding sites present on the respective protein. There are two different classes of binding sites.

1. *Independent, Identical binding sites*: If a protein possesses two independent but identical binding sites, then protein-ligand binding takes place as:

$$P + L \rightleftharpoons PL \quad (\text{binding constant}, K_1) \tag{7.51}$$

$$P + 2L \rightleftharpoons PL_2 \quad (\text{binding constant}, K_2) \tag{7.52}$$

Things to remember here is that
- If these binding sites are independent, binding of one ligand does not affect the other.
- For identical binding sites, $K_1 = K_2$.

If these binding sites of protein are independent, then one-to-one interaction takes place and simple addition will give the fraction of bound protein. If f is fraction or moles of ligand bound per mole of the macromolecule, then

$$f = \frac{n[L]}{K_d + [L]} = \frac{n[L]/K_d}{1 + [L]/K_d} \tag{7.53}$$

where n is number of binding sites and K_d is the dissociation constant of the protein-ligand complex. On rearranging this relation, we get

$$\frac{f}{[L]} = \frac{n}{K_d} - \frac{f}{K_d} \tag{7.54}$$

A plot of $\frac{f}{[L]}$ versus f is linear with slope equal to $-(1/K_d)$ and intercept equal to (n/K_d).

2. *Two classes of independent binding sites*: If there are two nonidentical binding sites on the protein (P) for a ligand.
The ligand binds with protein at the binding site a as follows

$$P_a + L \rightleftharpoons P_aL,$$

with $K_a = \frac{[P_a][L]}{[P_aL]}$ as dissociation constant of complex P_aL.
Ligand binds to the same protein at binding site b as follows
$P_b + L \rightleftharpoons P_bL$, with $K_b = \frac{[P_b][L]}{[P_bL]}$ as dissociation constant of the complex P_bL
The overall protein-ligand binding is represented as
$P_{ab} + 2L \rightleftharpoons P_{ab}L_2$, with $K_d = \frac{[P][L]^2}{[P_{ab}L_2]}$ as dissociation constant of complex $P_{ab}L_2$,

If both binding sites of the protein are independent, then

$$K_d = \frac{[P][L]^2}{[P_{ab}L_2]} = K_aK_b \tag{7.55}$$

In these cases, the fraction of bound protein is given as

$$f = \frac{[P_aL] + [P_bL] + 2[P_{ab}L_2]}{[P] + [P_aL] + [P_bL] + [P_{ab}L_2]} \tag{7.56}$$

The fraction of bound protein in terms of dissociation constant is expressed as

$$f = \frac{[P][L]/K_a + [P][L]/K_b + 2[P][L]/K_aK_b}{[P] + [P][L]/K_a + [P][L]/K_b + 2[P][L]/K_aK_b} \tag{7.57}$$

On rearranging, we get

$$f = \frac{[L]/K_a + [L]/K_b + 2[L]/K_aK_b}{(1 + [L]/K_a)(1 + [L]/K_b)} \tag{7.58}$$

and

$$f = \frac{[L]/K_a}{(1 + [L]/K_a)} + \frac{[L]/K_b}{(1 + [L]/K_b)} \tag{7.59}$$

If protein has n_1 independent sites with a dissociation constant K_1 and n_2 independent sites with dissociation constant K_2, then fraction of bound protein (f) is given by

$$\frac{f}{[L]} = \frac{n_1/K_1}{1 + [L]/K_1} + \frac{n_2/K_2}{1 + [L]/K_2} \tag{7.60}$$

7.6.2.4 Determination of Binding Constant from Fluorescence Quenching Data Analysis

Using the fluorescence quenching data, the binding constant of a protein-ligand complex can be determined using the Stern-Volmer equation. The binding constant can be obtained by the analysis of the quenching of fluorescence of ligand by the protein or vice versa. The Stern-Volmer equation is the relation that relates the fluorescence intensities in the presence as well as the absence of a quencher with the quencher concentration. It has been extensively used to describe both static and dynamic quenching phenomena (Lakowicz 2013). The Stern-Volmer equation is:

$$\frac{F_0}{F} = 1 + K_{sv}[Q] = 1 + k_q\tau_0[Q] \tag{7.61}$$

where F_0 and F are the fluorescence intensities of fluorophore present in protein in the absence and in the presence of quencher, respectively. $[Q]$ denotes the concentration of quencher, K_{SV} is Stern-Volmer quenching constant, k_q is the bimolecular quenching rate constant, and τ_0 is the lifetime of the fluorophore in the absence of quencher. K_{SV} can be equivalent to the association constant (K_a) only in case of static quenching. Figure 7.4 shows a typical Stern-Volmer plot for quenching of a protein fluorescence, where the slope of the plot determines the quenching constant (K_{SV}) for the respective system.

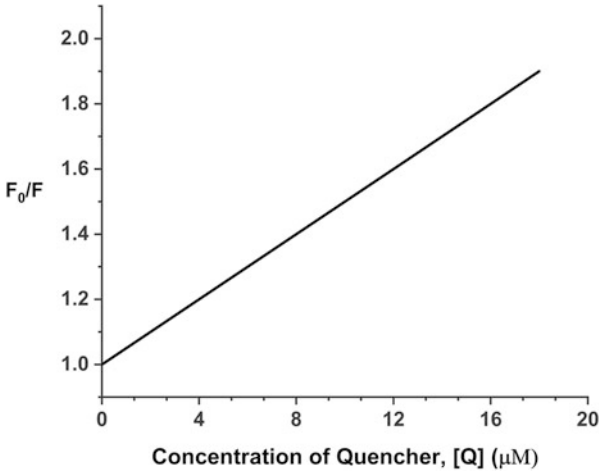


Fig. 7.4 Stern-Volmer plot for quenching of protein fluorescence

In the case where there is heterogeneous accessibilities of fluorophores, Stern-Volmer equation can be modified for determination of the fraction of bound protein as (Bourassa et al. 2013):

$$\frac{F_0}{F_0 - F} = \frac{1}{fK[Q]} + \frac{1}{f} \tag{7.62}$$

where F_0 and F depict the fluorescence intensities of protein fluorophore in the absence and the presence of quencher, respectively. $[Q]$ is the concentration of quencher, K is quenching constant, and f is the accessible fluorophore to the quencher. In this case, also, the plot of $F_0/(F_0 - F)$ vs. $[Q]$ is linear and intercept of the curve gives the fraction of bound protein while using both slope and intercept the quenching constant can be determined (Fig. 7.5).

For a quenching reaction between a protein and a quencher:

$$P + nQ \rightleftharpoons PQ_n$$

where P and Q stand for protein fluorophore and quencher, respectively. The binding constant (K_b) and the number of binding sites (n) can be calculated using another modified Stern-Volmer equation (Lakowicz 2013; Bourassa et al. 2013):

$$K_b = \frac{[PQ_n]}{[P][Q]^n} \tag{7.63}$$

We know that at any time of reaction, the total protein concentration is sum of the free protein and the bound protein, i.e.,

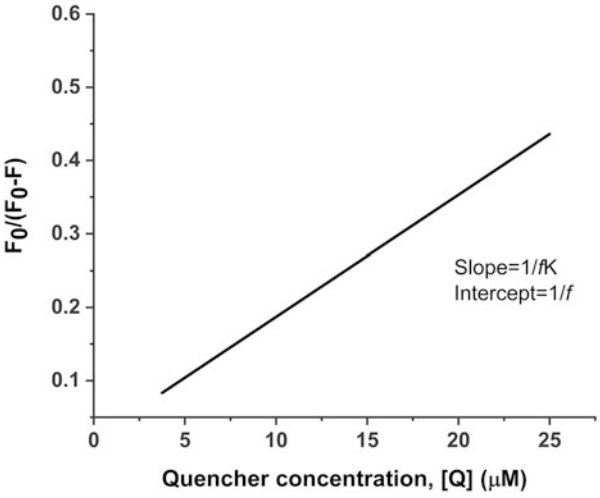


Fig. 7.5 Modified Stern-Volmer plot for calculation of fraction of bound protein

$$P_0 = P + PQ_n \tag{7.64}$$

Therefore, the binding constant can be expressed as

$$K_b = \frac{[P_0] - [P]}{[P][Q]^n} \tag{7.65}$$

or

$$\frac{[P_0]}{[P]} = 1 + K_b[Q]^n \tag{7.66}$$

If F_0 and F represents the fluorescence intensity of the protein fluorophore in free and bound state, respectively, then modified Stern-Volmer equation can be represented as:

$$\frac{[F_0]}{[F]} = 1 + K_b[Q]^n \tag{7.67}$$

Taking logarithm, we get

$$\log \frac{[F_0 - F]}{[F]} = \log K_b + n \log [Q] \tag{7.68}$$

If we plot $\log \frac{[F_0 - F]}{[F]}$ versus $\log[Q]$, a straight line will be obtained, the slope gives the number of binding sites while intercept provides the binding constant.

Using this relation, the binding constant and binding ratio of the protein and ligand were estimated for a number of complexes. Bhatia et al. (2015) showed that the curcumin binds with SOD1 in a 1:1 ratio and the binding constant was found to be $3.01 \times 10^4\ M^{-1}$. Also, Sharma and Pancholi (2014) showed that the fluorescence of BSA is quenched on its binding with olmesartranmedoxomil (OLM) and its metabolite (OL) via the static mechanism. The binding constant of BSA with OLM and OL was found to be 0.109 μM and 0.491 μM, respectively. Also, they found that the high affinity binding site available on BSA for OLM and OL were 2 and 1, respectively.

7.6.2.5 Other Methods to Calculate the K_d (Nonlinear Fitting of the Data)

There are also other methodologies to calculate the dissociation constant for a protein-ligand complex. If the data is not linear, then the nonlinear fitting is required. However, this method is more accurate for the determination of dissociation constant even if the assumption $[L] \approx L_T$ does not hold good. As we discussed in the previous section, the dissociation constant is expressed as:

$$K_d = \frac{[P][L]}{[PL]}$$

On rearranging this, we get

$$[P][L] - K_d[PL] = 0 \tag{7.69}$$

We know that at any time of reaction, $[P] = P_T - [PL]$ and $[L] = L_T - [PL]$. Therefore,

$$(P_T - [PL])(L_T - [PL]) - K_d[PL] = 0 \tag{7.70}$$

On rearranging we get,

$$[PL]^2 - (P_T + L_T + K_d)[PL] + P_T L_T = 0 \tag{7.71}$$

The solution of this quadratic equation is given as:

$$[PL] = \frac{-b \pm \sqrt{b^2 - 4P_T L_T}}{2} \tag{7.72}$$

where

$$b = -(P_T + L_T + K_d) \tag{7.73}$$

As discussed above, there are two cases: one where the signal is from the protein-ligand complex only and other where the protein in both free as well as bound form contributes to the resulting signal.

Case 1: when the signal is from protein complex only. The ratio of the concentration of protein complex to total protein can be expressed in terms of the signal as:

$$\frac{[PL]}{P_T} = \frac{S_T}{S_\infty} \tag{7.74}$$

or

$$\frac{S_T}{S_\infty} \times P_T = [PL] \tag{7.75}$$

Dissociation constant (K_d) can be calculated by nonlinear fitting of this equation.

Case 2: When the signal is from protein in free as well as bound form. In this case the ratio of protein complex to protein can be determined as:

$$\frac{[PL]}{P_T} = \frac{S - S_0}{S_\infty - S_0} \tag{7.76}$$

$$\frac{S - S_0}{S_\infty - S_0} \times P_T = [PL] \tag{7.77}$$

where S_0 denotes the initial signal, S denotes the signal at any time t, and S_∞ denotes the signal upon completion of the reaction.

For example, Deep et al. (2005) showed the nonlinear analysis of NMR data for the calculation of binding constant of a cytochrome c and cytochrome b_5 complex using chemical shift perturbation. They compared the chemical shift (δ) value for free form of cyt c with its complex form with cyt b_5. The equation used for the calculation of the binding constant (K_a) of the complex was:

$$\frac{\Delta\delta}{\Delta\delta_\infty} = -B - \frac{\sqrt{B^2 - 4ab}}{2a} \tag{7.78}$$

where

$$B = -(a + b + 1/K_a)$$

and a: total concentration of cyt c

b: total concentration of cyt b_5

$\Delta\delta$: change in chemical shift of cyt c on addition of cyt b_5

$\Delta\delta_\infty$: estimated change in chemical shift for the state when $b/a \rightarrow \infty$

From this chemical shift perturbation method, they found that the complex of cyt c and cyt b_5 is in fast exchange and binds in 1:1 stoichiometry. The binding constant for the complex was found to be $(4 \pm 3) \times 10^5\ M^{-1}$.

7.7 Calculation of Enthalpy/Entropy of Interaction

The enthalpy change of binding or dissociation can be calculated from the measurement of binding constant (K_a) or dissociation constant (K_d) or free energy change of protein-ligand interaction done at different temperatures. The van't Hoff analysis of the same is as follows:

$$\Delta H_d = \frac{d(\Delta G_d/T)}{d(1/T)} = -R\frac{d(\ln K_d)}{d(1/T)} \tag{7.79}$$

7.8 NMR as a Tool to Measure Thermodynamic Parameters of Binding

NMR is another useful analytical technique to study protein-ligand interactions. Generally, chemical shift mapping is employed to identify and study the binding area of the complex. The thermodynamic parameters of the complex formation like free energy of the binding and binding constant can be determined by the perturbation in a chemical shift as discussed above. NMR technique is preferred over other spectroscopic techniques because of its capability of measuring the occupancies of individual binding sites. Due to this, the microscopic binding constant of various complexes can be easily determined. This technique depicts the motion of protein molecules upon binding with the interacting ligand. The temperature-dependent NMR study of the complexes (in terms of free energy change) can determine the entropy and enthalpy components of the binding parameters.

The conformational entropy change for the protein-ligand complex formation can be determined using NMR study by measuring relaxation parameters like longitudinal relaxation rate (R_1), transverse relaxation rate (R_2), and the steady-state heteronuclear NOE. These relaxation parameters are defined as follows:

$$R_1 = \left(d^2/4\right)[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \tag{7.80}$$

$$R_2 = \left(d^2/8\right)[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \left(c^2/6\right)[4J(0) + 3J(\omega_N)] + R_{ex} \tag{7.81}$$

$$\text{NOE} = 1 + \left(d^2/4R_1\right)(\gamma_H/\gamma_N)[6J(\omega_H + \omega_N) - 6J(\omega_H - \omega_N)] \tag{7.82}$$

where d is the constant which defines the strength of the ^{15}N–H dipolar coupling and given as:

$$d = \frac{\mu_0 h \gamma_N \gamma_H \langle r_{NH}^{-3} \rangle}{8\pi^2} \tag{7.83}$$

and c is a constant given as

$$c = \frac{\omega_N \Delta\sigma}{\sqrt{3}} \tag{7.84}$$

μ_0 = permeability of free space,
h = Planck's constant,
γ_H = gyromagnetic ratio of ^1H
γ_N = gyromagnetic ratio of ^{15}N
r_{N-H} = bond distance of N–H bond = 1.02 Å
ω_H = Larmor frequency of ^1H
ω_N = Larmor frequency of ^{15}N
$\Delta\sigma = (\sigma_{\parallel} - \sigma_{\perp})$ = difference between parallel and perpendicular components of the axially symmetric ^{15}N chemical shift also called chemical shift anisotropy = −160 ppm
and, $J(\omega_i)$ = spectral density corresponding to *i*th nuclei

Here, the primary objective is to calculate different spectral densities and thereby deduce the physical aspects of the dynamics of the protein complexes. As the number of unknown parameters exceeds the available data, it is very difficult to extract the spectral densities directly using these three Eqs. (7.80)–(7.82). The most acceptable approach to determine the spectral density is model-free formalism.

7.8.1 Model-Free Formalism

Lipari-Szabo model-free formalism is the method, which does not include the assumptions regarding the physical model describing the molecular motions. In this methodology, only the internal dynamics are required, which is not correlated with the global tumbling of the biological system under study. According to this approach, spectral density at any Larmor frequency is determined using order parameters (S) and correlation times (τ_e) of the internal motions on fast as well as the slow time scale. If the overall and internal motions are assumed to be independent, then the isotropic correlation function with rotational correlation time (τ_m) will be defined as:

$$C(t) = e^{-t/\tau_m} C_I(t) \tag{7.85}$$

where $C_I(t)$ is the correlation function for the internal motions and given as:

$$C_I(t) = S^2 + \left(1 - S_f^2\right) e^{-t/\tau_{e,f}} + \left(S_f^2 - S^2\right) e^{-t/\tau_{e,s}} \tag{7.86}$$

where τ_m = isotropic rotational correlation time of the molecule, $\tau_{e,f}$ = effective correlation time of internal motion on fast time scale, $\tau_{e,s}$ = effective correlation time of internal motion on slow time scale, S = total generalized order parameter and given as $S^2 = S_f^2 S_s^2$, S_f = order parameters for the internal motions on fast time scale, and S_s = order parameters for the internal motions on slow time scale

Therefore, the corresponding spectral density function will be given as:

$$J(\omega) = \left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S_f^2)\tau_f}{1+(\omega\tau_f)^2} + \frac{(S_f^2-S^2)\tau_s}{1+(\omega\tau_s)^2}\right] \tag{7.87}$$

with

$$\tau_f = \frac{\tau_{e,f}\tau_m}{(\tau_{e,f}+\tau_m)} \text{ and } \tau_s = \frac{\tau_{e,s}\tau_m}{(\tau_{e,s}+\tau_m)} \tag{7.88}$$

Clore et al. (1990) showed that if the effective correlation time on fast scale (τ_f) is more than 10 ps, then Eq. (7.87) reduces to

$$J(\omega) = \left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(S_f^2-S^2)\tau_s}{1+(\omega\tau_s)^2}\right] \tag{7.89}$$

They examined the effect of different values of τ_f (0–50 ps) on the best fit values of τ_s, S_f^2, S_s^2, and S^2 and found that the most appropriate values of relaxation parameters can be obtained using Eq. (7.87) if τ_f is less than 50 ps. If its value exceeds beyond 50 ps, then Eq. (7.87) does not hold good for the evaluation of different relaxation parameters.

Lipari-Szabo model-free formalism involves stepwise calculation. In the first step, global rotational correlation time is calculated by excluding the residues (a) for which internal motions are very fast (low R_2/R_1 ratio and low NOE), (b) which are experiencing conformational exchange at a slow time scale. Now, the fitting of the iterative trimmed R_2/R_1 value ratio to an isotropic tensor will estimate the global rotational correlation time of the molecule. Then the data fit to five different models of the local motion for each residue to get the information about various order parameters, correlation times and rate of exchange.

7.8.2 Order Parameters (S^2)

The squared order parameter S^2 gives the angular amplitude of the reorientations of a bond. Model-free formalism is used to calculate the order parameters for each N–H or C–H bond vector. An order parameter of a vector gives an idea of the mobility of that vector. It is zero for the unrestricted bond motions and unity for a completely rigid vector. In general, its value lies around 0.85 for the ordered region and 0.4–0.6 for the disordered region. Order parameters can be obtained by analyzing the relaxation parameters, T_1, T_2, and NOE of the protein. Change of order parameter in going from free form to complex provides the information regarding the changes in mobility of the bond vector of particular residue. The mobility of the bond vector of a residue decreases with an increase in the value of the order parameter (S^2).

7.8.3 Order Parameter and Conformational Entropy

The change in backbone conformational entropy on the binding is related to change in backbone order parameters of residues (k) of protein in free (A) and bound form (B) through the equation:

$$\Delta S_{\mathrm{AB}} = R\sum_{k} \ln\left[\frac{1 - S_{B,k}^2}{1 - S_{A,k}^2}\right] \tag{7.90}$$

In a similar manner, Li and Bruschweiler proposed an equation to relate the change in side-chain entropy with an order parameter (Li and Bruschweiler 2009):

$$\Delta S_{AB} = R\sum_{m} C_m \sum_{n} \left(S_{B,n}^2 - S_{A,n}^2\right) \tag{7.91}$$

where C_m is a function of the number of side-chain dihedral angles in a given residue type and fit parameters and sums run over all residues n of type m.

Later, Sharp et al. gave the relation for extracting the total change in conformational entropy of a protein resulting from binding of the ligand from the NMR relaxation data (Sharp et al. 2015).

$$\frac{\Delta S}{k_{\mathrm{B}}} = 0.88 N_{\mathrm{res}} \left\langle \delta \ln\left(1 - O_{\mathrm{NH}}^2\right)\right\rangle - 0.56 N_{\chi} \left\langle \delta O_{\mathrm{axis}}^2 \right\rangle \tag{7.92}$$

where N_{res} = total number of residues, N_{χ} = total number of the side-chain chi angles, $\left\langle \delta O_{\mathrm{axis}}^2 \right\rangle$ = measured average change in methyl axis order parameter, $\left\langle \delta \ln\left(1 - O_{\mathrm{NH}}^2\right)\right\rangle$ = measured average change in amide order parameter

As it was seen in the Eq. (7.73), the configurational entropy also depends upon N_{χ} which signifies that the changes in order parameters (O_{axis}^2) account for the entropy changes from all side chains.

7.9 Computational Approach to Calculate the Thermodynamic Parameters of Binding

A computational approach to determine the protein interactions has great potential since this can provide insight not only about the contribution of different interactions but also the contribution of conformational and solvation entropic free energy.

7.9.1 From Changes in Solvent-Accessible Surface Area

The structural information of a protein-ligand binding process can be related to its thermodynamic parameters. Spolar and Record (1994) developed an empirical

relationship, which relates the heat capacity change with the change in solvent-accessible surface area (ΔSASA) upon binding.

$$\Delta C_{\mathrm{p}}\left(\mathrm{cal\ K^{-1}mol^{-1}}\right) = -0.33\left(\mathrm{cal\ K^{-1}mol^{-1}\AA^{-2}}\right)\ \Delta \mathrm{ASA_{np}}\left(\mathrm{\AA^2}\right) + 0.16\left(\mathrm{cal\ K^{-1}mol^{-1}\AA^{-2}}\right)\ \Delta \mathrm{ASA_{p}}\left(\mathrm{\AA^2}\right) \tag{7.93}$$

This relation was further modified by Deep and Ahluwalia (2002), in which disulfide crosslinks were also considered.

An estimation of enthalpy change can be done using accessible surface area. The enthalpy change at reference temperature $60^{\circ}C$ is equal to

$$\Delta H_{\mathrm{bind}}\left(60^{\circ}\mathrm{C}\right) = \Delta h_{\mathrm{ap}} \cdot \Delta \mathrm{ASA} + \Delta h_{\mathrm{p}} \cdot \Delta \mathrm{ASA_p} \tag{7.94}$$

The enthalpy change at other temperature can be calculated using

$$\Delta H_{\mathrm{bind}}(T) = \Delta H_{\mathrm{bind}}\left(60^{\circ}\mathrm{C}\right) + \Delta C_{\mathrm{p}}(T - 333.15) \tag{7.95}$$

The calculated enthalpy of binding using these equations is reasonable for protein-protein and protein-peptide binding; however, correlation is not good for protein binding with the small ligand.

Experimental enthalpy obtained from ITC will have three components:

$$\Delta H_{\mathrm{exp}} = \Delta H_{\mathrm{bind}} + \Delta H_{\mathrm{conf}} + \Delta H_{\mathrm{prot}} \tag{7.96}$$

ΔH_{bind} is enthalpy of binding, ΔH_{conf} is heat released or absorbed due to conformational change, ΔH_{prot} is the enthalpy change due to protonation/deprotonation during binding.

The total entropy change also has three components

$$\Delta S_{\mathrm{tot}} = \Delta S_{\mathrm{solv}} + \Delta S_{\mathrm{conf}} + \Delta S_{\mathrm{rt}} \tag{7.97}$$

ΔS_{solv} is entropy change due to expulsion of water from protein surface on binding, ΔS_{conf} is the change in entropy due to change in side-chain conformational entropy associated with protein and ligand on binding, and ΔS_{rt} is change in entropy due to loss of rotational and translational degree of freedom upon binding (Murphy et al. 1995).

The contribution of side chain to ΔS_{conf} can be calculated using changes in accessible surface of side chain on binding

$$\Delta S_{\mathrm{conf}} = \sum_i \frac{\Delta \mathrm{ASA}_{\mathrm{sc},i}}{\mathrm{ASA}_{\mathrm{AXA},i}} \Delta S_{\mathrm{bu}\rightarrow\mathrm{ex},i} \tag{7.98}$$

$\Delta \mathrm{ASA}_{\mathrm{sc},\,i}$ is the change in the accessible surface area of side chain of a particular amino acid i, $\mathrm{ASA}_{\mathrm{AXA},\,i}$ is the accessible surface area of the side chain ith amino acid

(X) in ALA-X-ALA tripeptide. $\Delta S_{\mathrm{bu} \rightarrow \mathrm{ex},\, i}$ is entropy change of side chain on being exposed from buried, the value is different for side chains of different amino acid (Baker and Murphy 1998).

7.9.2 Molecular Docking

Molecular docking is a commonly used method for calculating the binding free energy of protein-ligand complexes. In comparison to other methods like MD and MC, docking is cheaper and very fast processing tool. With computational power in hand today, it is possible to screen hundreds of molecules in an hour. Thus, docking serves as a powerful tool in the field of drug discovery. Over the last two decades, a large number of docking software packages have been emerged for the calculation of binding free energy. AutoDock, GOLD, DOCK, FlexX, and Glide are some of the most commonly used docking software (Morris et al. 2009; Jones et al. 1997; Kuntz et al. 1982; Rarey et al. 1996; Friesner et al. 2004). Different software uses different algorithms, e.g., GOLD uses the genetic algorithm (Mitchell 1998), AutoDock uses a hybrid local search GA known as the Lamarckian genetic algorithm (LGA) which performs better than GA for Autodock, similarly FlexX employed an incremental construction (IC) algorithm (Kramer et al. 1999).

In general, protein-ligand docking has two key steps, namely, search algorithms and scoring functions (Du et al. 2016). The search algorithm scans various poses of a ligand inside a given target whereas scoring functions arrange these generated poses according to their binding affinities with the target. There are two types of treatments available for protein and ligand in the search algorithm: rigid body and flexible body treatment. Various search algorithms use a combination of these two treatments, e.g., ZDock (Chen et al. 2003), MDock (Pierce et al. 2005), and older version of DOCK use rigid body treatment; in this both protein and ligand structures are taken as rigid bodies; a very less degrees of translational and rotational freedom is available to ligand. These search algorithms are the fastest and simplest ones.

The second type of search algorithm treats protein as a rigid body and ligand has all the freedom to translate and rotate inside the protein cavities to get the best fit. The new version of DOCK utilizes this search algorithm. The third and more exact search algorithms consider the flexibilities of both protein and ligand, but these are the most compute intensive search algorithms. It is now possible to mark a certain region of protein as flexible and the rest of it can be kept rigid; this approach is faster and gives appreciable results.

After search algorithm then comes the importance of scoring functions which are used to calculate the binding free energy of protein-ligand complex after docking. For an accurate calculation of binding free energies, the scoring function should account solvent interactions and entropic contributions as well but due to complex calculations and computational limitations, it is very difficult.

Scoring functions can be categorized into three classes: force-field based, empirical, and knowledge based. Force-field is a set of parameters that has bonding and nonbonding terms for a molecule. These parameters are generally derived from

experiments and ab initio calculations. Force-field-based algorithms gave information about intermolecular non-covalent interactions, i.e., enthalpic contribution, while entropic contribution has not been accounted for in this. Empirical scoring functions are simplified as compared to force-field-based scoring functions. These functions utilize regression and machine learning methods to parameterize any interaction as favorable or unfavorable. The applicability of these methods depends on the accurate energy description.

The knowledge-based scoring functions consider those interactions energetically favorable that occur a greater number of times over some random interactions. These energetically favorable interactions only contribute to the binding energy. The knowledge-based algorithms are simpler and faster than force-field-based functions and not likely to suffer from overfitting or overtraining as compared to empirical functions. Although docking methods are used widely but these are limited to only solved protein structures and due to availability of less degree of freedom docking cannot include the contribution of entropic changes and solvation effects are also not accounted.

7.9.3 Force-Field-Based Methods

For quantitative estimation of binding free energy protein's dynamics has to be considered and for that various force-field-based methods can be used, e.g., molecular dynamics and Monte Carlo simulations. These methods are computed intensive but provide the intrinsic entropic contributions towards the total free energy change for protein-ligand complexes.

7.9.3.1 Free Energy Perturbation (FEP) Method

The free energy perturbation (Zwanzig 1954) method was first introduced by Robert W. Zwanzig (1954). FEP method is based on the alchemical FE calculations in which free energy is calculated via various physical processes involving the formation of intermediates. According to these calculations, free energy difference for going from state A to state B is obtained from the well-known Zwanzig equation:

$$\Delta G(\mathrm{A} \rightarrow \mathrm{B}) = G_{\mathrm{B}} - G_{\mathrm{A}} = -k_{\mathrm{B}}T\left\langle \mathrm{e}^{\left(-\frac{E_{\mathrm{B}}-E_{\mathrm{A}}}{k_{\mathrm{B}}T}\right)} \right\rangle \tag{7.99}$$

where T is temperature, k_{B} is Boltzmann's constant. Triangle brackets represent ensemble averages generated through simulations. E_{A} and E_{B} represent potential functions of the state A and B. Convergence of FEP calculations is ensured when only a small difference exists between the two states, so it is better to break down the perturbation into a series of smaller windows which can be computed independently. Total free energy is obtained by summing the individual free energies from separate steps. FEP method can be used to study the host-guest binding energetics, pk_{a} predictions, for studying solvent effects on reactions and for enzymatic reactions.

7.9.3.2 Thermodynamic Integration (TI) Method

TI is another method, which is based on alchemical calculations (Straatsma and Berendsen 1988). In TI, the free energy change is calculated for going from state A to state B along a thermodynamic path connecting the two states and for this an order parameter "λ" is defined. The value of λ ranges from 0 to 1. The potential energy as a function of λ can be written as:

$$E(\lambda) = E_{\mathrm{A}} + \lambda(E_{\mathrm{B}} - E_{\mathrm{A}}) \tag{7.100}$$

The binding free energy can be obtained from MD or MC by taking the ensemble average of the potential energy derivative with respect to λ and integrating it from $\lambda = 0$ to $\lambda = 1$. TI is the unidirectional transformation from A to B, unlike FEP where the transformation from B to A is also possible.

$$\Delta G(\mathrm{A} \rightarrow \mathrm{B}) = \int_0^1 \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{7.101}$$

With the help of thermodynamic integration method, we can study the relative free energy of solvation of two species, calculate the relative binding constant for complex formation, and can also study the relative stability of mutant proteins.

7.9.3.3 Lambda Dynamics Simulations

λ dynamics simulations (Knight and Brooks 2009) utilize path sampling calculations in which a physically possible path is followed for the conformational transition in order to calculate the binding free energies. In λ dynamics, simulations order parameter λ is treated as a dynamic variable with fictitious mass and is varied throughout the simulation. Therefore, it is possible to estimate free energy from a single simulation, unlike FEP and TI where multiple simulations have to be run. λ dynamics simulations overcome the convergence problem present in FEP calculations. Total energy change is obtained by summing over involved intermediate states along the λ variable.

$$\Delta G = \sum_{i=0}^{n-1} \Delta G(\lambda_i \rightarrow \lambda_{i+1}) \tag{7.102}$$

The value of λ ranges from 0 to 1 and it is divided into n small increments, $\Delta\lambda$. Alternatively, the free energy can be calculated along the reaction coordinate. One of the major advantages of using λ dynamics over traditional FEP and TI method is that the thermodynamic properties of multiple molecules can be computed in a single simulation. Alternate binding modes of a ligand can be explored in a single simulation.

7.9.3.4 Linear Interaction Energy (LIE) Method

Linear interaction energy method (Wang et al. 1999) is an end point method in which only free and bound states are simulated for binding free energy calculations. Due to

sampling and convergence problems, alchemical methods are not very useful for structure-based drug design. Therefore, a fast and accurate semi-empirical method, linear interaction energy approximation was introduced. The reason of the faster performance of LIE is that in-between states of free and bound forms are not sampled in this method. LIE method considers that change in interaction energy upon any alteration in ligand surrounding directly affects binding free energy. The binding free energy can be evaluated from the following equation:

$$\Delta G = \alpha\left(\langle E_{\text{elec}}\rangle_{\text{bound}} - \langle E_{\text{elec}}\rangle_{\text{free}}\right) + \beta\left(\langle E_{\text{vdw}}\rangle_{\text{bound}} - \langle E_{\text{vdw}}\rangle_{\text{free}}\right) \tag{7.103}$$

$E_{\text{vdw,bound}}$ and $E_{\text{elec,bound}}$ are VW and electrostatic interaction energies between the ligand and the solvated protein and $E_{\text{vdw,free}}$ and $E_{\text{elec,free}}$ are the VW and electrostatic interaction energies between ligand and water. α and β are two empirical parameters. α originated from first-order approximation of electrostatic contribution to binding free energy and β is derived empirically. β acts as a scaling factor for the difference in VW interaction energies between bound and free states; therefore its value is influenced by binding site. Success of the LIE method depends on the values of parameters α and β; this reliance may limit the application, predictive power, and efficiency of the method.

7.9.3.5 Molecular-Mechanics-Poisson-Boltzmann Surface Area (MM-PBSA) Method

Another free energy calculation method which is based on the end point calculations is known as the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) (Kollman et al. 2000). This is a widely used method for protein-ligand binding free energy estimation. Free energy can be calculated from:

$$\Delta G = \Delta G(\text{vacuum}) + (\Delta G_{P-L}(\text{solv}) - \Delta G_P(\text{solv}) - \Delta G_L(\text{solv})) \tag{7.104}$$

where ΔG (vacuum) is vacuum binding free energy, $\Delta G_{\text{P–L}}$, ΔG_{P}, and ΔG_{L} are solvation free energies of protein-ligand complex, free protein, and free ligand, respectively. Solvation free energies can be calculated either from three independent simulations of complex, free ligand and free protein or this information can be extracted from the single trajectory of protein-ligand complex by removing protein for ΔG_{L} and ligand for ΔG_{P}. MM-PBSA is an explicit solvent model. After removing the explicit solvent molecules ΔG (vacuum) can be obtained from the following equation:

$$\Delta G\,(\text{vacuum}) = \langle E_{P-L}\rangle - \langle E_P\rangle - \langle E_L\rangle - T\Delta S_{\text{solute}} \tag{7.105}$$

where $<E_{P-L}>$, $<E_P>$, and $<E_L>$ are the Boltzmann average of the potential energy of protein-ligand complex, free protein, and free ligand, respectively. ΔS_{Solute} is the change in entropy upon binding which can be estimated with the help of statistical thermodynamic models. The solvation energy ΔG (solv) is computed from PBSA implicit solvent model. The interactions between biomolecules such as protein-protein, protein-peptide, and protein-nucleic acid (Wang et al. 2019) can

be anticipated using MM-PBSA method. Owning to the speed and accuracy of this method, it has applications in many aspects of drug designing.

7.9.3.6 Hybrid Quantum Chemical/Mechanical (QM/MM) Method

For calculation of thermodynamic properties like free energy change upon ligand binding, an accurate force field and converged phase space sampling both are important. For this quantum mechanical methods are often used but their application is limited to finite size systems. A system with a large number of degrees of freedom like proteins cannot be treated with high accuracy QM approach due to very high computational cost. Therefore, hybrid QM/MM methods have been developed by combining the two levels of theory. In hybrid QM/MM, a small number of degrees of freedom, which are thought to be the most important, will be treated with high level theory, i.e., QM and rest of the system are modeled according to simplified descriptor, i.e., MM. Therefore, a balance between precision and affordable complexity can be achieved with hybrid QM/MM.

In combined QM/MM methods, generally, the system is divided into two subsystems: a small QM subsystem containing the active site and a large MM subsystem consisting of the remaining molecular system. The sum of different interaction terms gives the total potential energy of a QM/MM system (Hu and Yang 2009):

$$\begin{aligned} E_{\frac{\mathrm{QM}}{\mathrm{MM}}}(\mathrm{QM},\mathrm{MM}) = {} & E_{\mathrm{QM}}(\mathrm{QM},\mathrm{MM}) + E_{\frac{\mathrm{QM}}{\mathrm{MM}},\mathrm{elec.}}(\mathrm{QM},\mathrm{MM}) \\ & + E_{\frac{\mathrm{QM}}{\mathrm{MM}},\mathrm{nucl.}}(\mathrm{QM},\mathrm{MM}) + E_{\frac{\mathrm{QM}}{\mathrm{MM}},\mathrm{vdw}}(\mathrm{QM},\mathrm{MM}) \\ & + E_{\frac{\mathrm{QM}}{\mathrm{MM}},\mathrm{cov}}(\mathrm{QM},\mathrm{MM}) + E_{\mathrm{MM}}(\mathrm{MM}) \end{aligned} \tag{7.106}$$

where E_{QM} is QM internal energy, $E_{\mathrm{QM/MM,elec.}}$ is the electrostatic interaction energy between QM electrons and MM subsystem, $E_{\mathrm{QM/MM,nucl.}}$ is electrostatic interaction energy between the QM nuclei and MM subsystem, $E_{\mathrm{QM/MM,vdw}}$ is VW energy, and $E_{\mathrm{QM/MM,cov}}$ is covalent interaction energy between two subsystems. E_{MM} is purely the MM interaction energy of the MM subsystem.

Hybrid QM/MM methods can be combined with other FE calculation methods like thermodynamic integration (TI) method, free energy perturbation method (FEP), and umbrella sampling method. However, due to the high computational demand for treating the QM subsystem, their application is still limited. Several enzymatic reactions and electron transfer reactions have been studied with hybrid QM/MM methodology. In spite of its inherent strength, the MD methods suffer from two major bottlenecks. The results are dependent on the force field used and therefore the use of an appropriate force field is critical. Due to the increased computational cost of MD simulations, the binding is investigated on a relatively short time scale resulting in inadequate sampling and hence biased towards starting structure. Some of the methods used to overcome the problem of samplings are replica exchange MD and metadynamics.

References

Ahluwalia U, Nayeem SM, Deep S (2011) The non-native conformations of cytochrome c in sodium dodecyl sulfate and their modulation by ATP. Eur Biophys J 40(3):259–271

Arango Duque G, Descoteaux A (2014) Macrophage cytokines: involvement in immunity and infectious diseases. Front Immunol 5:491

Baker BM, Murphy KP (1998) Prediction of binding energetics from structure using empirical parameterization. Methods Enzymol 295:294–315

Bezerra GA, Dobrovetsky E, Viertlmayr R, Dong A, Binter A, Abramić M et al (2012) Entropy-driven binding of opioid peptides induces a large domain motion in human dipeptidyl peptidase III. Proc Natl Acad Sci U S A 109(17):6525

Bhatia NK, Srivastava A, Katyal N, Jain N, Khan MA, Kundu B et al (2015) Curcumin binds to the pre-fibrillar aggregates of Cu/Zn superoxide dismutase (SOD1) and alters its amyloidogenic pathway resulting in reduced cytotoxicity. Biochim Biophys Acta 1854(5):426–436

Biswas T, Tsodikov OV (2010) An easy-to-use tool for planning and modeling a calorimetric titration. Anal Biochem 406(1):91–93

Bourassa P, Bariyanga J, Tajmir-Riahi HA (2013) Binding sites of resveratrol, genistein, and curcumin with milk alpha- and beta-caseins. J Phys Chem B 117(5):1287–1295

Broecker J, Vargas C, Keller S (2011) Revisiting the optimal c value for isothermal titration calorimetry. Anal Biochem 418(2):307–309

Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. Proteins 52 (1):80–87

Chodera JD, Mobley DL (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. Annu Rev Biophys 42:121–142

Clayton AH, Hanley QS, Arndt-Jovin DJ, Subramaniam V, Jovin TM (2002) Dynamic fluorescence anisotropy imaging microscopy in the frequency domain (rFLIM). Biophys J 83(3):1631–1649

Clegg RM, Holub O, Gohlke C (2003) Fluorescence lifetime-resolved imaging: measuring lifetimes in an image. Methods Enzymol 360:509–542

Clore GM, Driscoll PC, Wingfield PT, Gronenborn AM (1990) Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy. Biochemistry 29(32):7387–7401

Cohen AS (1994a) Clinical aspects of amyloidosis, including related proteins and central nervous system amyloid. Curr Opin Rheumatol 6(1):68–77

Cohen AS (1994b) Proteins of the systemic amyloidoses. Curr Opin Rheumatol 6(1):55–67

Cooper A (1999) Thermodynamic analysis of biomolecular interactions. Curr Opin Chem Biol 3 (5):557–563

Cooper A, Johnson CM, Lakey JH, Nollmann M (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. Biophys Chem 93(2–3):215–230

Dallas SL, Sivakumar P, Jones CJ, Chen Q, Peters DM, Mosher DF et al (2005) Fibronectin regulates latent transforming growth factor-beta (TGF beta) by controlling matrix assembly of latent TGF beta-binding protein-1. J Biol Chem 280(19):18871–18880

Deep S, Ahluwalia JC (2002) Heat capacity of folding of proteins corrected for disulfide cross-links. Biophys Chem 97(1):73–77

Deep S, Im S-C, Zuiderweg ER, Waskell L (2005) Characterization and calculation of a cytochrome c− cytochrome b 5 complex using NMR data. Biochemistry 44(31):10654–10668

Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P et al (2016) Insights into protein-ligand interactions: mechanisms, models, and methods. Int J Mol Sci 17(2):E144

Fersht AR (1972) Conformational equilibria in α-and δ-chymotrypsin: the energetics and importance of the salt bridge. J Mol Biol 64(2):497–509

Fielding L (2003) NMR methods for the determination of protein-ligand dissociation constants. Curr Top Med Chem 3(1):39–53

Fowler A, Swift D, Longman E, Acornley A, Hemsley P, Murray D et al (2002) An evaluation of fluorescence polarization and lifetime discriminated polarization for high throughput screening of serine/threonine kinases. Anal Biochem 308(2):223–231

Frasca V (2016) Biophysical characterization of antibodies with isothermal titration calorimetry. J Appl Bioanal 2(3):90

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47(7):1739–1749

Gautier I, Tramier M, Durieux C, Coppey J, Pansu RB, Nicolas JC et al (2001) Homo-FRET microscopy in living cells to measure monomer-dimer transition of GFP-tagged proteins. Biophys J 80(6):3000–3008

Harpur AG, Wouters FS, Bastiaens PI (2001) Imaging FRET between spectrally similar GFP molecules in single cells. Nat Biotechnol 19(2):167–169

Hart PJ, Deep S, Taylor AB, Shu Z, Hinck CS, Hinck AP (2002) Crystal structure of the human TbetaR2 ectodomain--TGF-beta3 complex. Nat Struct Biol 9(3):203–208

Hill TL (1986) An introduction to statistical thermodynamics. Courier Corporation, New York

Hill RL, Brew K (1975) Lactose synthetase. Adv Enzymol Relat Areas Mol Biol 43:411–490

Holdgate GA (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. BioTechniques 31(1):164–166

Hu H, Yang W (2009) Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes. Theochem 898(1–3):17–30

Jameson DM, Croney JC, Moens PD (2003) Fluorescence: basic concepts, practical aspects, and some anecdotes. Methods Enzymol 360:1–43

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267(3):727–748

Knight JL, Brooks CL 3rd. (2009) Lambda-dynamics free energy simulation methods. J Comput Chem 30(11):1692–1700

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L et al (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33(12):889–897

Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins 37(2):228–241

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161(2):269–288

Ladbury JE (2010) Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. Biochem Soc Trans 38(4):888–893

Lafont V, Armstrong AA, Ohtaka H, Kiso Y, Mario Amzel L, Freire E (2007) Compensating enthalpic and entropic changes hinder binding affinity optimization. Chem Biol Drug Des 69 (6):413–422

Lakowicz JR (2013) Principles of fluorescence spectroscopy. Springer Science & Business Media, New York

Le VH, Buscaglia R, Chaires JB, Lewis EA (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. Anal Biochem 434(2):233–241

Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11(5):560–566

Li DW, Bruschweiler R (2009) A dictionary for protein side-chain entropies from NMR order parameters. J Am Chem Soc 131(21):7226–7227

MacRaild CA, Daranas AH, Bronowska A, Homans SW (2007) Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein. J Mol Biol 368(3):822–832

Mitchell M (1998) An introduction to genetic algorithms. MIT Press, Cambridge

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30 (16):2785–2791

Murphy KP, Freire E, Paterson Y (1995) Configurational effects in antibody-antigen interactions studied by microcalorimetry. Proteins 21(2):83–90

Nadi MM, Ashrafi Kooshk MR, Mansouri K, Ghadami SA, Amani M, Ghobadi S et al (2015) Comparative spectroscopic studies on curcumin stabilization by association to bovine serum albumin and casein: a perspective on drug-delivery application. Int J Food Prop 18(3):638–659

Pearlman DA, Connelly PR (1995) Determination of the differential effects of hydrogen bonding and water release on the binding of FK506 to native and Tyr82→ Phe82 FKBP-12 proteins using free energy simulations. J Mol Biol 248(3):696–717

Pepys MB, Hawkins PN, Booth DR, Vigushin DM, Tennent GA, Soutar AK et al (1993) Human lysozyme gene mutations cause hereditary systemic amyloidosis. Nature 362(6420):553–557

Phizicky EM, Fields S (1995) Protein-protein interactions: methods for detection and analysis. Microbiol Rev 59(1):94–123

Pierce MM, Raman CS, Nall BT (1999) Isothermal titration calorimetry of protein-protein interactions. Methods 19(2):213–221

Pierce B, Tong W, Weng Z (2005) M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. Bioinformatics 21(8):1472–1478

Porpaczy Z, Sumegi B, Alkonyi I (1983) Association between the alpha-ketoglutarate dehydrogenase complex and succinate thiokinase. Biochim Biophys Acta 749(2):172–179

Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. J Mol Biol 261(3):470–489

Savage G, Morrison S. Trypsin inhibitors. 2003

Sharma RN, Pancholi SS (2014) Protein binding interaction study of olmesartan medoxomil and its metabolite olmesartan by fluorescence spectroscopy. Int J Pharm Pharm Sci 6(2):726–729

Sharp KA, O'Brien E, Kasinath V, Wand AJ (2015) On the relationship between NMR-derived amide order parameters and protein backbone entropy changes. Proteins 83(5):922–930

Sipe JD (1992) Amyloidosis. Annu Rev Biochem 61(1):947–975

Sommerfeld A (1964) Thermodynamics and statistical mechanics—lectures on theoretical physics, vol 5. Academic, New York

Spolar RS, Record MT Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science 263(5148):777–784

Sprang SR (2016) Activation of G proteins by GTP and the mechanism of Gα-catalyzed GTP hydrolysis. Biopolymers 105(8):449–462

Stockmann H, Bronowska A, Syme NR, Thompson GS, Kalverda AP, Warriner SL et al (2008) Residual ligand entropy in the binding of p-substituted benzenesulfonamide ligands to bovine carbonic anhydrase II. J Am Chem Soc 130(37):12420–12426

Straatsma TP, Berendsen HJC (1988) Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. J Chem Phys 89(9):5876–5886

Subramaniam V, Hanley QS, Clayton AH, Jovin TM (2003) Photophysics of green and red fluorescent proteins: implications for quantitative microscopy. Methods Enzymol 360:178–201

Tellinghuisen J (2008) Isothermal titration calorimetry at very low c. Anal Biochem 373 (2):395–397

Turconi S, Shea K, Ashman S, Fantom K, Earnshaw DL, Bingham RP et al (2001) Real experiences of uHTS: a prototypic 1536-well fluorescence anisotropy-based uHTS screen and application of well-level quality control procedures. J Biomol Screen 6(5):275–290

Turnbull WB, Daranas AH (2003) On the value of c: can low affinity systems be studied by isothermal titration calorimetry? J Am Chem Soc 125(48):14859–14866

Tzeng SR, Kalodimos CG (2012) Protein activity regulation by conformational entropy. Nature 488 (7410):236–240

Ventre I, Ledgham F, Prima V, Lazdunski A, Foglino M, Sturgis JN (2003) Dimerization of the quorum sensing regulator RhlR: development of a method using EGFP fluorescence anisotropy. Mol Microbiol 48(1):187–198

Vershon AK, Liao SM, McClure WR, Sauer RT (1987) Interaction of the bacteriophage P22 Arc repressor with operator DNA. J Mol Biol 195(2):323–331

Walker JE (1978) In: Hoffmann E (ed) Proteins: structure, function and industrial applications. Pergamon Press, Oxford

Wang W, Wang J, Kollman PA (1999) What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? Proteins 34(3):395–402

Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH et al (2019) End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. Chem Rev 119(16):9478–9508

Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179(1):131–137

Yan Y, Marriott G (2003a) Analysis of protein interactions using fluorescence technologies. Curr Opin Chem Biol 7(5):635–640

Yan Y, Marriott G (2003b) Fluorescence resonance energy transfer imaging microscopy and fluorescence polarization imaging microscopy. Methods Enzymol 360:561–580

Yin F, Cao R, Goddard A, Zhang Y, Oldfield E (2006) Enthalpy versus entropy-driven binding of bisphosphonates to farnesyl diphosphate synthase. J Am Chem Soc 128(11):3524–3525

Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. J Chem Phys 22(8):1420–1426

Synergistic Effects of Hydration Sites in Protein Stability: A Theoretical Water Thermodynamics Approach

8

Jayashree Biswal, Prajisha Jayaprakash, Raghu Rangaswamy, and Jeyaraman Jeyakanthan

Abstract

The hydrating water molecules play a crucial role in enhancing the structural stability, protein folding, and competition with the ligand for its occupancy at the binding site. This chapter describes the importance of hydrating water molecules in biomolecular processes explained via thermodynamic approaches. The focus is on understanding the change of magnitude of hydrating water molecules present at the surface of the protein and their shift to the bulk water and protein environment. Due to significant inhomogeneity in water density, polarity, and mobility assisted by intramolecular interactions such as hydrophobic, electrostatic forces, and hydrogen-bonding network, the thermodynamic information are hard to obtain through experiments. Computational elucidation of water displacements by ligand molecules is characterized by the entropic and enthalpic compensation where the inhibition of mobility of the water molecule and solvent environment sets a limit on the entropic contribution and ligand reactivity. Most importantly, computational simulations can provide a better understanding of the most intriguing aspects of hydration water molecules. Water is an influential determinant for protein–protein interactions, protein–nucleic acid interactions, and other enzymes-assisted or cofactor-assisted interactions that are driven by hydrophilic forces, whereas the hydrophobic effects facilitate the binding energy of the ligand.

J. Biswal · P. Jayaprakash · J. Jeyakanthan (✉)
Structural Biology and Bio-Computing Lab, Department of Bioinformatics, Science Block, Alagappa University, Karaikudi, Tamil Nadu, India

R. Rangaswamy
Structural Biology and Bio-Computing Lab, Department of Bioinformatics, Science Block, Alagappa University, Karaikudi, Tamil Nadu, India

Schrödinger India, Bengaluru, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_8

Keywords

Protein hydration · Enthalpy and entropy compensation · Hydrophilic and hydrophobic · Water displacement and thermodynamics

8.1 Introduction

Drug designer uses proteins as a therapeutic option that becomes challenging without knowing the nature and the conformational characteristics of the specific protein (Lagasse et al. 2017). Proteins are represented by four forms of structures (primary structure, secondary structure, tertiary structure, and quaternary structure). There are 20 naturally occurring amino acids, grouped into ten polar and ten nonpolar amino acids based on their side chain features (Kangueane and Nilofer 2018). Analytical methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy are used to determine the structural stability of the proteins. Protein stability can be disrupted by various factors such as external stress factors including temperature, pH, removal of water, presence of hydrophobic surfaces, metal ions, etc. (Thomas 2020; Tripathi 2013). Water plays a crucial role in protein folding and actively participates in the long range of water-mediated contacts through the hydrogen-bonded network. Similarly, adding water molecules will enhance protein folding and drug design strategies with high specificity (Rhee et al. 2004). Each water molecule acts as a magnet comprising three atoms (two hydrogen atoms that have a positive charge and one oxygen atom that has a negative charge). The oxygen atom of the nucleus (with eight positively charged protons) in the water molecule attracts electrons strongly than the hydrogen nuclei (with only one positively charged proton). The different part of the water molecule has a separate charge named as a polarity in which a particular part of the molecule is positively charged, and another part of the molecule is negatively charged.

Water molecules play an essential role in the stability, folding, dynamics, and function of the biomolecules (Levy and Onuchic 2004). The hydration forces determine the stabilization and packing of the protein structure. These hydrating water molecules involved in the formation of hydrogen bond networks. Hydration of water molecules close to the protein surface that exhibits dynamical properties from those of bulk are crucial for stabilizing the folded proteins (Bizzarri and Cannistraro 2002). Hydrating water fluctuations can affect the protein function, and disturb the protein dynamics factor, having consequent changes in the protein motions. A solvent is an active participant in energy associated with a protein (Ansari et al. 1992; Fenimore et al. 2002). Hydrophobic interactions contribute to protein stability and are very important for making protein stable and biologically active conformation (Zapadka et al. 2017). The energetics of protein folding and binding involves significant trade-offs between losses of protein–water interactions.

Nowadays, thermodynamic signatures have become popular and are considered as the most important criteria for the selection of potential drug molecules (Ladbury

et al. 2010). The only limitation from the thermodynamic signatures is their unpredictable motions in the local water structure. Due to the thermodynamic nature of the binding, water molecules have a dominant effect. Specific interactions were observed in water molecules for the promotion of protein–ligand, but the thermodynamic signals are not involved in the proximity of the ligand (Spyrakis et al. 2017). Water not only interacts with the protein surface, but it will also have interplay with the protein backbone and side chains (Levy and Onuchic 2004). Mutations affect the placement of the water molecules in the protein structures and disturb the main interactions of the water network that leads to destabilization (Covalt et al. 2001). Interactions observed in water-mediated network exhibit entropy-enthalpy compensations, which has a minor effect on affinity and significant impact on the thermodynamic signatures. The contributions of solvent–solvent interactions play a vital role in molecular recognition processes. Therefore, experimentally predicted binding enthalpy value is extracted from isothermal titration calorimetry and have a mix of several positive and negative thermodynamic signatures which strongly restrain the solvent changes (Ruhmann et al. 2015).

The major limitation and accuracy of molecular simulations depend on the relevant biological systems and the core force fields. A biological system consists of large molecules with several internal degrees of freedom and solvent molecules. Sampling of all conformations is a complicated process and inorder to overcome these issues, we can calculate the free energy perturbation (FEP) simulations, which give differences in estimated free energy between closely related systems, e.g., small incremental changes in ligands (Manzoni and Ryde 2018). The selection of appropriate force field is very challenging in the case of nonavailable experimental evidence.

The water in the hydrogen bond is neither weak nor too strong. Each water molecule forms two hydrogen bonds and are covalently attached to the oxygen of a water molecule (492.21 kJ mol^{-1}) (Boyarkin et al. 2013). To predict the location of a particular water molecule from the same crystal structure, different methods and their applied force fields can give huge different estimates on the position and conformation. A recent study has investigated the relationship between water models and the computed free energies in protein folding (Anandakrishnan et al. 2019). Computed protein folding free energy landscape obtained from water–water interactions seems to have the most significant discrepancies due to the long range of electrostatic interactions and not because of Van der Waals interactions (Kuffel 2017).

This chapter will review the identification of the location of water molecules and its thermodynamics properties (ΔH_{solv} and ΔS_{solv}), which are very important for protein stability and the drug discovery process. Protein hydrations are essential for the determination of protein 3D structures (Wüthrich et al. 1992). Most of the water molecules interact with the protein, and the resulting hydration structure is an integral part of the proteins. Due to the amino acid substitution on the protein surface, slight changes are observed in protein stability and hydration. There are various tools available to predict the hydration sites, and it will be explained in the subsequent section, which may help in defining the solvent sites through different

computational techniques applied on an explicitly solvated binding pocket (Mondal et al. 2014). This extensive thermodynamics study will give a significant insight on molecular design and drug discovery.

8.2 Water Molecules in Proteins

Water molecules present in the protein or between the proteins are linked by hydrogen bonds similar to the interface between a ligand and protein. When compared with the protein, protein–ligand complexes (one or two) with the presence of water has a higher B-factor (Lu et al. 2007). The term "buried" defines that a molecular entity's solvent-accessible surface area is $\leq 5\%$, and the B-factor of such buried water holding three or four bonds is lower than the structural average of the protein because of their location in different secondary structure elements. Significantly, low B-factor with buried water is conserved in the crystal structures of evolutionarily linked proteins (Takano et al. 2003).

Interestingly, the B-factor relation also holds vice versa: Protein atoms show lower B-factor when hydrogen bond is formed to buried water molecules exclusively, rather than hydrogen bonding to another protein atom. Similarly, the protein backbone amide nitrogen can form only one H-bond, while the carbonyl oxygen can form two simultaneously. Mostly, amino acid side chains buried in the protein core are hydrophobic, so these water molecules forms hydrogen bonds to the protein backbone. Hence, the hydrophobic forces are involved in the denaturation of the proteins and their temperature dependency (Van Dijk et al. 2015).

After the hydrophobic folding process, water uses protein to satisfy their requirement of hydrogen bond formation. Consequently, backbone hydrogen bonds formed with the water molecules and not with other protein residues are localized in regions that are neither involved in helical structure nor pleated sheet conformation (Finkelstein and Ptitysyn 2016). Waters are mostly observed in α-helices than β-sheets and often found in coil regions. The difference between α-helices and β-sheets for hosting water molecules is not because of the number of cavities or hydrophobicity but it is likely due to secondary structure flexibility and thus residence times (Gromiha 2000).

Buried polar side chains are flexible or evolutionarily optimized enough to find other protein atoms to bond and do not rely on water molecules. Protein–ligand interface consists of charged and polar amino acids such as Arg and Glu, responsible for the main hydrated side chains. Commonly hydrophobic residues such as Val, Leu, Ile, Phe, and Met have an interior buried polar side chain exposed to solvent (Malleshappa Gowder et al. 2014). Lys is a charged amino acid, frequently hydrated in rigid protein–ligand interfaces, but when it has a large cavity, obtain rotameric states (Perutz et al. 1965). Such characteristics are mostly found in disordered crystal structures that impart this disorder on nearby water molecules. The amino acids Tyr and Trp constituting aromatic moieties are more hydrating in nature.

In apo-protein crystal structures, among all 20 amino acids, Gly has the lowest hydration propensity because of its extensive flexibility and small size. This situation

is reversed in the rigid interfaces of protein–ligand complexes, where it has the highest main chain hydration propensity comparable to the Ser side chain as its backbone is accessible than the backbone of other amino acids (Pappas et al. 2012). On the other hand, the amide nitrogen of Pro does not form hydrogen bonds. Yet, it is frequently observed to be located close to buried water molecule due to its rigidness, B-factor will decrease. Interestingly, Pro in protein–ligand systems has the lowest hydration propensity when compared to other side chains (Rose and Wolfenden 1993). Gly and Pro behave inversely as Pro is a nonpolar amino acid that gets displaced upon ligand binding.

There is no cutoff kept for buried water molecules present in protein surface, but the overwhelming bulk is found at less than 3 Å, or about one protein atom depth. Nearly 60% of buried waters are "alone," 20% is as a cluster of two molecules, 8% as a cluster of three molecules, and so on. Each additional water molecule beyond the first forms about 1–1.5 hydrogen bonds to the protein, while the remaining interacts with the other water molecules (Maurer and Oostenbrink 2019). Generally, waters present in the core region will enhance the protein stability with the hydrating polar atoms (Ball 2008).

8.2.1 Protein Hydration: An Essential Factor for Various Biological Processes

As mentioned, water or protein hydration is vital to form and maintain the 3D structural and functional aspects of protein. The combined use of experimental (X-ray, neutron scattering and diffraction, NMR, terahertz spectroscopy) and computational (Molecular simulations) techniques have confirmed the essentiality of water in protein stability and consequent processes (Bellissent-Funel et al. 2016). Layers of hydrating water molecules determine the protein structure resolution and reliability. Proteins solved in a resolution of about 1.5–1.6 Å have a continuum of hydration layers with a mono-layer covering >1.5 moles H_2O mol^{-1} amino acid residue.

8.2.2 Role of Water Molecules in Protein Stability

Proteins are essential for biological functions and are polymers of alpha-amino-acids. Under normal pH, temperature, and ionic strength conditions, the native structure of protein assumes an ideal 3D fold. The proteins hold different functional groups in their side chains (amide, carboxylic, hydroxyl, thiol, aromatic). This specific native protein structure catalyzes and regulates reactions, transport substrates, code, and transcribes genetic information. It has been extensively appreciated that water molecules are essential for maintaining the structure, stability, dynamics, and function of biomolecules. In reality, the absence of water molecules may lead to a lack of functional activity. However, it is found in recent years that water has been prominently treated as an integral component of biomolecular

systems. Mixtures of experimental and computational studies acknowledge the active role of water in biomolecule structure, stability, and dynamics. Molecular dynamics (MD) simulations can balance the knowledge gained through experiments and explain the biomolecule and solvent as well as the time dependency of their dynamics (Karplus and McCammon 2002). The importance of water in structure, stability, and dynamics of proteins and nucleic acids is discussed below, besides the underlying molecular recognitions are also addressed.

8.2.3 Hydrating Water Around the Protein Surface

Water molecules interact with the surface groups by reorientation of both themselves and protein. In contrast, the other water molecules link this association to the rest of the bulk in an orderly manner to remain in active form (Zhong et al. 2011). The movement of the surface water molecules is due to the hydrogen bonding, and the dielectric constant is constrained, which is imposed by the protein (Seyedi and Matyushov 2018). The surface of the protein contains the electron acceptor groups than the donor, along with the excluded-volume effects, which is the cause for unoccupied hydrogen bonds in the region of the neighboring water molecules. The number of hydration sites in a protein depends on two factors, namely (a) conformational variability and (b) the freedom that the water has to hydrate the protein (Parsegian 2002). Protein conformations requiring greater hydration are facilitated by the compactness of waters containing many weak, delicate, or twisted hydrogen bonds, whereas the other form of conformations requires low-density water containing many strong intramolecular hydrogen bonds. The waters around the surface region are held captive by the basic amino acids, and those waters, when exposed to the bulk solvent through the groups present in that region, paves the way to greater flexibility and chain movement (Bandyopadhyay et al. 2005a, b). The core hydration layers around the protein surface are affected by the chemical composition of the amino acids and molecular interactions. Conversely, hydrating water molecules are slower in the presence of hydrophobic residues than the hydrophilic residues.

Adding water molecules can ease the structure prediction through the knowledge-based potential to an established Hamiltonian. The study shows that water not only induces protein folding and binding but also actively participates via water-mediated contacts. Water present at the surface of biological macromolecules defines a layer, which has been termed “biological water.” The closeness of the “biological water” at the protein surface distinctly exhibits dynamical properties from that of the bulk, also its residence times in the sub-nanosecond range.

8.2.4 Protein Folding–Hydrogen Bond Network–Water Clusters–Domain Motions

The hydrating water molecules influence the reactions and interactions of coenzymes and cofactors. Despite changes in the ionic strength, there seems to be a less impact on the protein charge distribution extended by the effects of hydrating waters (Virtanen et al. 2014). Second, the most important biological phenomena "protein folding" depends on the drive of the hydrophobic groups, which is usually clustered away from the protein surface and is controlled by the interactions with polar based residues and cooperative hydration (Huggins 2016). Also, the protein folding is characterized by the decrease in the β-sheet linking hydrogen bonds (Novak and Grdadolnik 2017). For a protein to be biologically active, it should form a 2D hydrogen-bonded network spanning the protein surface connecting all the surface hydrogen-bonded water clusters (Oleinikova et al. 2005; Smolin et al. 2005). Changes in hydration sites are necessary to account for the collective domain motions, with water molecules corresponding to structural transitions through movements of the protein cavity observed through alteration in the formation of hydrogen bonds. Besides this, interior protein motions are combined through the involvement of hydrogen-bonded water bound in the protein structure that facilitate through the distribution of information between the functionally important regions of the protein.

8.2.5 Electrostatic Forces and Buried Waters

Intramolecular forces contribute significantly to protein structure and stability, and the forces are more influential in α-helices rather than β-sheets (Nakasako 2004). Water molecules can help to bridge different peptide links by catalyzing the interaction between the carbonyl oxygen atoms and amide protons, forming more interactions that stabilize protein–ligand and protein–protein interfaces (Cao and Bowie 2014). Internal water molecules enable the folding of proteins and are cast away from the hydrophobic central core (Ikura et al. 2004). MD simulations will facilitate the computation of buried waters, and multi-water bridges (Cheung et al. 2002), and these buried waters form structurally important hydrogen-bonded linkages. Usually, the first hydration shell around the protein surface has high proton transfer, well-resolved hydration sites, and organized hydrogen bond interactions with large net dipole fields (Pradhan et al. 2019). Hydrogen bonds formed by these hydration waters around the surface have longer lifetimes than the bulk water (Yokomizo et al. 2005). The organization of the waters present in the nearby proteins expands its electrostatic visibility to visiting ligands (Chakraborty et al. 2007). Moreover, the protein folding nature can be determined by the electrostatic effect of the water directed towards specific hydrogen bonding, which extends into the bulk from the surface of hydrogen-bonding to the amino acid side chains (Hildebrandt et al. 2007).

8.2.6 Role of Thermodynamics in Protein Stability

Temperature and pressure play a significant role in protein unfolding when the temperature of the system increases, the entropy also increases, and proteins adapt themselves in folded form. The major role of hydration on protein folding can be studied by MD simulation, umbrella sampling, coarse-grained models, and knowledge-based structure prediction models. These calculations showed that despite the hydrophobic disruption in the protein chain, the water molecules enrich the folding process by forming intermediates with the protein backbone. Entropy is considered being the most critical parameter in understanding the stability of the proteins. Thermodynamic studies help to predict the protein conformational changes under environmental conditions and thermodynamic parameters useful in analyzing protein stability. The thermodynamic properties (enthalpy, entropy, and free energy) always have a standard state while it changes during folding and unfolding of proteins, which aids in understanding the protein stability (Gummadi 2003).

Entropy is a mathematical concept that describes the distribution of energy within a system. Free energy is a thermodynamic function that relates enthalpy and entropy to spontaneity and equilibrium constants. Hence there are theoretical approaches such as WaterMap (Sect. 8.3.10) that will help in locating the hydration sites and thermodynamic properties of water molecules. This facilitates solvate protein-binding sites that offer rich physical insights into the properties of the pocket and the hydrophobic forces driving the binding of small molecules.

8.2.7 Thermodynamics Properties in the Stability of Molecular Interactions

Water is an essential part of enzymes, which penetrates and determines how molecules move, bind to each other, and solvate charges. For example, the proton transfer takes place inside of proteins carried out by the Grotthuss mechanism, and it requires a group of hydrogen bonds involving internal water molecules and protein amino acid residues. Consecutively, water molecules can assist the enzymes catalytic reactions by behaving as a temporary proton donor/acceptor. Hence, to understand the function, it is essential to know about the location of internal water molecules available in the protein structures. To find the location of the internal water molecules, it is best to observe while they are forming hydrogen bonds, but due to its dynamic nature, the crystallographic prediction of the hydrating water molecules becomes less feasible. Only crystal structures with high resolution can be able to capture a few water molecules with certainty per protein (Kubinyi 2001).

Non-covalent interactions play a significant role in deducing the structure stability, and specific Thermodynamic and Spectroscopic studies are used to identify and compute the non-covalent interactions. Due to the enthalpic and entropic contribution, there is a difference in free energy between native and unfolded states of a protein. Numerous forces bring small and differing contributions to protein stability. Few recent experimental techniques can elucidate the structure and denatured states

of the protein. The ignorance of the structural details of unfolded proteins makes it difficult to analyze all data quantitatively. Electrostatic interactions, particularly the salt bridges and cation interactions, also the Van der Waals interactions, play essential roles in defining the stability of proteins. The *hydrophobic effect, hydrogen bonds, and water molecules* are important contributors to protein structure and stability.

Frank and Evans explained the low solubility of nonpolar variety in the water at the molecular level. In the late 1950s, Kauzmann introduces the concept of "hydrophobicity," which explains the protein folding complexity (Baldwin 2014). Water is very fundamental in protein folding because of hydrophobic attractions. Water form clusters around nonpolar groups (hydrophobic hydration), which lead to slight decrease in entropy of the system, and when they are released from the hydrophobic surfaces, there is a gain in entropy contributing to protein stability. The hydrophobic "collapse" of the protein is necessarily accompanied and explored by hydrogen bond formation between favorable functional groups (Fernández et al. 2003). For biological phenomena like protein folding water molecules are essential to define the "hydrophobic interactions." Hydrogen bonds will have a positive contribution towards protein stabilization and balance between entropy and enthalpy terms. Regardless of the little involvement build towards protein stability via hydrogen bonds, if an intermolecular hydrogen bond in a protein is broken or deleted without the possibility of forming a compensating hydrogen bond to solvent, the protein will lose its structure and get destabilized.

Hydrogen bonding is nothing but dipole-dipole interaction between molecules, e.g., the hydrogen bond between water molecules; O atom is attracted to an H atom in the second molecule. The strength of a hydrogen bond is around 4–50 kJ mol^{-1}. This is, however, not necessarily the amount of energy that the hydrogen bond contributes to the stabilization of a folded protein. In the unfolded state, potential hydrogen-bonding partners in the polypeptide chain are satisfied by hydrogen bonds to water. When the protein folds, the hydrogen bonds are broken, some are replaced by intra-protein hydrogen bonds and the entropy of the solvent increases. The balance between the entropy and enthalpy terms is close, but hydrogen bonds contribute positively to protein stability. Despite the small contribution made to protein stability by hydrogen bonds, if an intermolecular hydrogen bond in a protein is broken or deleted without the possibility of forming a compensating H-bond to solvent, the protein will be destabilized and can lose its structure (Ragone 2001).

Apart from conferring stability to protein's structure, hydrating waters take part in most of the protein–protein (Lehmkuhler et al. 2017), protein–DNA ((a) Rodier et al. 2005 (b) Lo Conte et al. 1999), and protein–ligand ((a) Lu et al. 2007 (b) Panigrahi and Desiraju 2007; Reddy et al. 2001) interactions and aid in the molecular recognition and both the binding thermodynamics and kinetics (Bodnarchuk 2016) properties. Enthalpy contribution is characterized by its mobility in the displacement process, and complete arrest sets a limit on its involvement (Maurer and Oostenbrink 2019). Usually, protein–protein and protein–DNA binding surfaces have hydrophilic residues, and their interaction happens with the help of water molecules. The comparative parameters between the binding cavity and the ligand are hydrophobic.

The binding energy attributes to changes in the free energy when two hydrophobic surfaces not in contact can be seen in hydrophobic effects (Chaplin 2008).

8.2.8 Displacement of Water Molecules in the Protein Cavity and Its Associated Thermodynamics

The kinase inhibitors have a specific recognition mechanism following the water occupancy in the binding pocket (Maurer and Oostenbrink 2019). Attributes of water, such as its smaller size, polarity, conformational flexibility, interaction strength, and directions, directly contribute to elasticity and reversibility. The main force for binding relies on the displacement of the unstable waters, and the dependent energetic gain through favorable relative binding energy is feasible through molecular rearrangement of the hydration water molecules (Snyder et al. 2014; Levinson and Boxer 2014; Lim et al. 2012; Jana and Bandyopadhyay 2012).

The thermodynamics concept of ligand binding is influenced by the relationship between enthalpy and entropy contributions of the binding event. Protein–ligand interactions involve attractive forces and hydration effects. Accurately positioned polar groups will make a way to specific interactions (Hydrogen bonds, salt-bridges, polar-polar interactions, and non-classical interactions) such as whole mediated halogen bonding results in enthalpy gain. To use this enthalpy, binding partners should be in optimal orientation, since the binding energy is highly sensitive to both the distance and the angle of the interacting atoms. The feature of interactions and the associated binding thermodynamics profile impact selectivity against off-targets (Tarcsay and Keserű 2015). Enthalpically optimized compounds possess carefully positioned ligand-binding site atom pairs to achieve the desired gain in binding enthalpy. Due to the improper orientation of the ligand, the desired protein interactions cannot yield the enthalpic contribution to binding free energy. In contrast, entropically optimized compounds have less positional constraints, and desolvation of the polar moieties can result in entropy gain due to the lower dependence from the binding environment. The compounds have, therefore, a higher propensity to form attractive interactions with off-targets. Through, the analysis of the thermodynamic properties, the structural and functional characteristics of the protein can be easily elucidated by the occupancy of high-energy hydration sites. Somewhere high-energy hydration sites are mostly localized near hydrophilic protein motifs (Olsson et al. 2008). Furthermore, there was no significant correlation between the hydration site-free energy and the solvent-accessible surface area of the site. Besides, the distribution of high-energy hydration sites on the protein surface can identify the location of binding sites and that binding sites of druggable targets have a higher density of thermodynamically unstable hydration sites.

The water molecules observed in crystal structures are less stable on average than bulk water due to the high degree of spatial localization resulting in loss of entropy. These findings must help to a better understanding of the water characteristics at the surface of proteins and lead to insights into the structure-based drug design efforts. Nonpolar ligand groups assist displacement of waters from the binding site due to

the decrease in the interference of the previously bound water with the protein's internal hydrogen-bonding and improved bulk hydrogen-bonding (Snyder et al. 2014).

8.3 Solvent Mapping Tools

In this section, we will discuss a few commercial and freely available tools and software that can predict the water hydration sites. The list of open-source software is listed in Table 8.1.

8.3.1 WatAA

This tool is newly developed which is known as atlas of amino acid hydration in proteins (Černy et al. 2017) that helps in exploring the synergies between data mining and ab-initio calculations using Turbo-Mole v6.4 program (Ahlrichs et al. 1989) with the DFT-D/RI-TPSS/TZVP method and calculating the interaction

Table 8.1 Different concepts involved in the identification of hydration sites around the protein

Hydration site tool	Theory	References
WatAA: Web-based atlas of amino acid hydration in proteins	Data mining study of protein crystal structures	Černy et al. (2017)
WATCLUST: A protein solvation structure analysis tool	Inhomogeneous fluid solvation theory	Lopez et al. (2015)
SZMAP (solvent-Zap-mapping)	Semi-explicit solvent mapping approach uses semi-continuum	Santa Fe (2013)
3D-RISM: Three-dimensional reference interaction site model	Molecular solvation theory	Fusani et al. (2018)
ProBiS H_2O	Local structure alignment algorithm	Jukic et al. (2017)
WaterFLAP: Fingerprints for ligands and proteins	Continuum solvent method and GRID molecular interaction fields (MIFs)	Baroni et al. (2007)
WAP: Water analysis package	Computing engine CGI/PERL	Shanthi et al. (2002)
Consolv	Knowledge-based, prediction-based hybrid K nearest algorithm	Raymer et al. (1997)
WaterScore	Multivariate logistic models	Garcia-Sosa and Mancera (2003)
WaterMap	Inhomogeneous solvation theory	Abel et al. (2008)
$DOWSER^{++}$	Semi-empirical modification program using energy-based MD simulation	Morozenko and Stuchebrukhov (2016)
WaterDock	Quality threshold (QT) algorithm and semi-empirical approaches	Ross et al. (2012)
WATsite	Quality threshold algorithm	Hu and Lill (2014)

energies (E_{int}) of biomolecular fragments; the calculations were compared with the computationally intensive benchmark CCSD (T)/CBS method (Jurecka et al. 2007). The available data in the atlas was taken from two sources, such as experimental data and ab-initio quantum mechanics calculations. Validation of water replacement using electron density maps in crystallographic refinement helps in locating water molecules mediating protein–ligand interactions and MD simulations. The quantum mechanics calculations and validation can be performed by optimization and stabilizing the water position of each hydration site, by providing hydrogen atom positions and quantifying the interaction energy. A non-redundant set of 2818 high-resolution protein crystal structures are collected from protein data bank (PDB) and classified according to the secondary structure and X1 rotameric state. Data mining analysis provides the yielded statistical data on each amino acid residue hydration site. Hydration sites are positioned near local energy minima, and interaction energies were calculated that helps in assessing the water molecule hydration sites individually. Fourier averaging was also performed for the water densities and displacement. The displacement term is the spatial distance between the position of the crystal-derived water molecule and its optimized position (Schneider et al. 1993).

8.3.2 WATCLUST

This tool helps in identifying the specific and freely available water sites (WS). MD simulation (explicit water) can be performed using the VMD program. By using trajectory files, the WATCLUST plug-in can be used by the option "Extensions > Analysis > WATCLUST." This plug-in helps in determining the WS and subgroup of residues where the WS will be calculated. The program computes: (a) water finding probability (WEP), (b) R90, (c) WS-protein mean interaction energy ($<E\omega\rho>$), (d) WS water mean interaction energy with respect to bulk (ΔE_{int}), and finally (e) excess rotational (S_r) and translational (S_t) entropies (Lopez et al. 2015).

This method is mostly applicable for the identification of WS in protein–ligand binding sites, presence of hydrophilic hot-spots in the protein–protein interface and exclusively aids in the identification of water structures, waters in the ion channels, and arrangement of catalytic waters. The disadvantage of this method is that it cannot be applied to the regions highly hydrophobic.

8.3.3 SZMAP (Solvent-Zap-Mapping)

SZMAP and GamePlan, developed by OpenEye scientific available at Santa Fe (2013). The gamePlan is used for analyzing the water sites; further running of SZMAP provides the coordinates of the protein–ligand binding site and analyzing the results, respectively (Grant et al. 2001). SZMAP is a semi-explicit solvent mapping (Tanger and Pitzer 1989; Rashin and Bukatin 1991) approach that uses the semi-continuum model to study the thermodynamics properties of water in the

protein–ligand system using both explicit and implicit solvent modeling methods and address the displacement of the water-binding site which either increases or decreases the binding affinity. Specific water orientations were analyzed using classical statistical mechanics and sampling. SZMAP calculates ΔG, ΔH, and $T\Delta S$, which differentiate the explicit probe and the continuum water. A positive ΔG indicates that the continuum model estimates the cost of displacing water at that position. This will help in comparing the water probe to the ligand atom that displaces it. The difference in free energy between standard and uncharged water is negative, where standard water is more favored and positive. SZMAP can calculate stabilization energies from the neutral difference values as reaction energy:

$$(\text{holo} - \text{complex} + \text{bulk} - \text{solvent}) - (\text{apo} - \text{pocket} + \text{free} - \text{ligand})$$

where bulk-solvent is defined as 0 kcal mol^{-1}.

SZMAP predicts very low B-factors where water molecule has less entropy and is more buried. Therefore, the crystallographic waters occupy hydrophilic positions on an SZMAP n_ddG contour map. Higher burial terms (near one) had lower B-factors. In addition, it could predict water conservation across a series of protein–ligand complexes and also predict optimal directions for deriving ligands during lead optimization. The SZMAP neutral probe entropy difference (n_dTdS) is a single physics-based descriptor that provides an accurate prediction of water conservation in active sites of protein structures using a simple threshold-value model.

8.3.4 3D-RISM

3D reference interaction site model (3D-RISM) is developed by a chemical group and works on a genetic algorithm; besides, the local minima problem was neglected by the desirability function. For detecting the potential water sites, 3D-RISM uses a double filter procedure and the minimum threshold value for the density distribution. Gridpoint was set as default $g(r) > -5$, followed by spatial constraint applied in the center and radius of the grid. The density distribution from the 3D-RISM calculation is transformed to a population function by using the equation $P(T) = \rho_{\text{bulk}}\, V_{\text{voxel}}\, g(\vec{r})$, where ρ_{bulk} is the density of the bulk solvent, V_{voxel} is the volume of one voxel in the grid, and $g(\vec{r})$ is the density function. Following this, each water site is detected in the first phase of the program. Followed by scoring was defined as the weighted ratio of the number of incompatible water sites. Hence this tool will be helpful in prediction and calculation of solvent density distribution. Besides this, GA_{sol} is capable of finding the network of water molecules that best fits a particular 3D-RISM density distribution rapidly and accurately (Fusani et al. 2018).

8.3.5 ProBiS H_2O

ProBiS H_2O is a solvent mapping tool, which helps in the prediction of conserved water sites available in the protein data bank (PDB) by using a local structure alignment algorithm. It is an innovative approach, uses the existing experimental structures, and helps in the prediction of conserved water sites. This program is the first tool to perform local superimposition for the prediction of conserved water using DroP algorithm with Cambridge Structural Database (CSD) (Colin et al. 2016), a small molecule crystal database used for defining the water molecules interactions using geometric criteria, and Acqua Alta algorithm used to reproduce the water molecules interactions (Rossato et al. 2011). The density-based spatial clustering of applications with noise (3D-DBSCAN) algorithm was implemented in the scikit-learn machine learning Python library and clusters were defined as dense regions defined as the ε-neighborhood of a data point (p) with n or more data points (q). A dense region comprising a cluster is then calculated as:

ε-neighborhood of an object contains at least n objects $p\varepsilon$-neighborhood

$N\varepsilon\ (p)$:$\{q \backslash d(p,q) < \varepsilon\}$ (ε-neighborhood with objects (p,q) within a radius ε from an object)

Setting of ε to 0.9, this equal to a sphere with a radius of 0.9 Å. Then n parameter is increased iteratively from $n = 1$, indicating random water molecules, to $n = N$, where N is the number of superimposed chains of similar proteins.

The repetitively clustering calculation was done with an increase in the density until no more clusters are recognized. The ProBiS H_2O plug-in is useful in finding the conserved and water networks in protein and designing the drug molecule and plays a role in protein-structure stability (Jukic et al. 2017).

8.3.6 WaterFLAP

A new approach developed by molecular discovery for predicting binding site water molecules. It uses a continuum solvent method and Grid inhomogeneous solvation theory (GIST) Molecular Interaction Fields (MIFs). Water scoring was done using two new GRID fields: the combined hydrophobicity and lipophilicity (CRY) field for combined hydrophobicity and lipophilicity, entropic (ENTR) field to estimate the entropic character of a particular water molecule and OH_2 water enthalpy. In addition to this, water enthalpy prediction and various properties like structural, displaceable, and bulkiness can be assessed. GRID-binding sites in terms of their MIFs enable straight forward structure-based design. Water network creation can also be done using WaterFLAP; the water is placed on the most favorable positions (Baroni et al. 2007).

8.3.7 WAP

A web-based package used to calculate geometrical parameters between water molecules present in the protein and nucleic acid structure (Shanthi et al. 2002). There are two ways of implementing the program: (a) structure available from the PDB. This tool will provide information about the unit cell parameters, symmetry, and space group to the users. Distances are set to 2.5 Å (minimum) and 3.6 Å (maximum), followed by an angle set to default 0. (b) 3D atomic coordinates from the client machine. Here, the user has to provide input PDB files from the client machine. It displays the chains, metal ions, and inhibitor information. Protein angle and distance were calculated. This package gives information about all side chain/backbone polar atoms, protein, water, nitrogen, oxygen atoms.

8.3.8 Consolv

Consolv uses knowledgebase prediction, hybrid K nearest neighbors' classifier, and genetic algorithm for the identification of hydration sites (free and ligand-bound water molecules) in the free environment; the water molecule's crystallographic temperature factor, the number of hydrogen bonds between the water molecules and protein, density, and hydrophilicity of the neighboring protein atoms. The ability to predict water-mediated and polar interactions have an essential role in protein structure and function. A training set of 13 non-homologous proteins was used in the identification of conserved and displaced water molecules in the active site with 75% accuracy. Water molecules within 3.6 Å of protein surface atoms capable of making Van der Waal's contacts or hydrogen bonds to atoms in the protein are considered being first-shell waters (De Beer et al. 2010; Raymer et al. 1997).

8.3.9 WaterScore

WaterScore is a method developed by researchers (Garcia-Sosa and Mancera 2003) for identifying between bound and displaceable water molecules. The structural properties of water molecules can be performed using multivariate logistic statistical analysis. WaterScore uses multivariate logistic models (Glonek and McCullagh 1995) and develops structural features of water molecules present in the empty binding site of a protein. There is a chance of observing the same water molecules (bound) after ligand binding. In WaterScore, the generated model will be considered based on the B-factor, the solvent-contact surface area, the total hydrogen bond energy, and the number of protein atomic contacts. This consistent approach gives an idea that the addition of water molecules towards the binding site of protein transversely differs from the various biomolecular applications. Models will be verified by checking the water molecules that are placed in bounded form or displaced in the binding site by using a better resolution of 3D structure. Secondly, in the structure-based drug design, the methodology adopts sorting, analyzing, and

including the selection of water molecules in the protein structures. These criteria will help in the prediction of more accurate binding mode and energies for designing ligands. This tool also has several applications extended to de novo drug design and molecular docking and plays a significant role in ligand binding.

8.3.10 WaterMap

WaterMap software is developed by Schrödinger, based on inhomogeneous solvation theory (Lazaridis 1998). It shows the hydration site around the ligand-binding site. It can be used for enzymes such as GPCRs, bromodomains, nucleic acids, and protein–protein interfaces. Salient features of WaterMap (WaterMap, Schrödinger, LLC, New York, NY, 2020) includes identification of the water sites and thermodynamic properties present in the protein-binding sites and detailed examination of the thermodynamics binding comprising the free energy changes resulting from displacing water molecules in the active site (Fig. 8.1). It is represented by a sphere where there is a region of space, and the water molecules tend to aggregate. Therefore, each hydration site has several associated thermodynamic properties. WaterMap calculation is a multistep process; the first step is "simulation," where the system is simulated for 2 ns (300 K) with full explicit solvent and heavy restraints were applied. The second step is "Clustering," the position of water molecule will be clustered based on the clustering algorithm and water density, which can be measured for each position. The highest water location represents the hydration

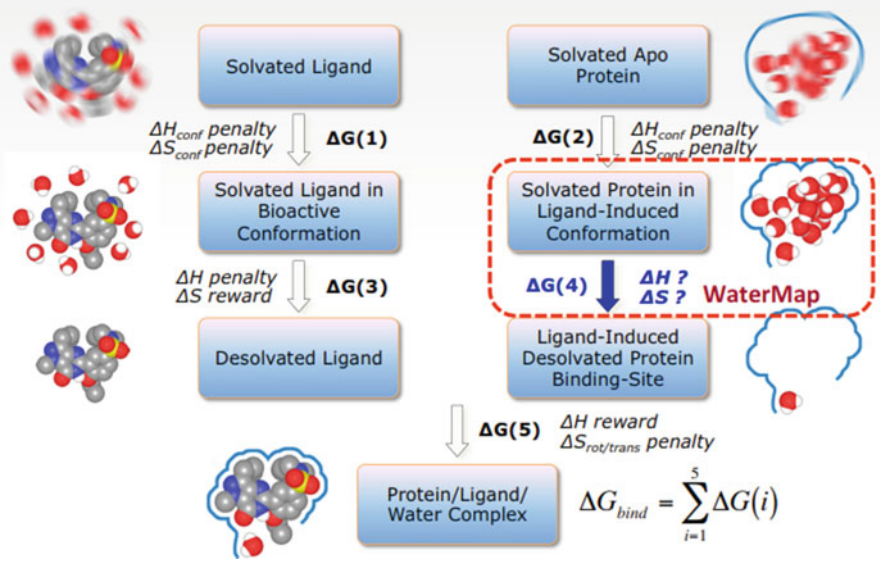


Fig. 8.1 Schematic representation of the thermodynamic decomposition of ligand/protein binding adapted from WaterMap Schrödinger (WaterMap 2020)

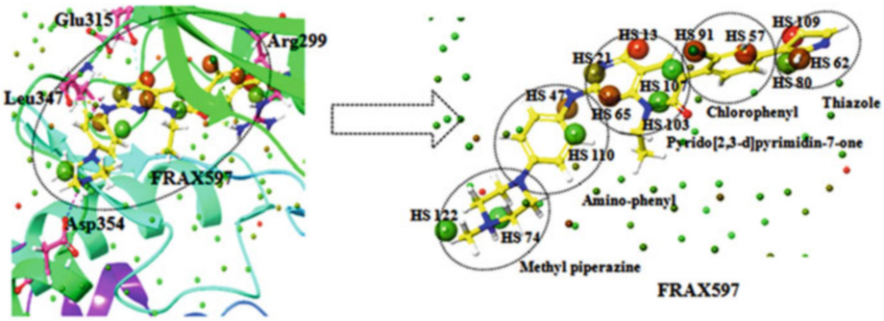


Fig. 8.2 Localization of computationally identified hydration sites overlapping with Pak1 crystal structure (PDB id: 4EQC) and its (A) co-crystal ligand FRAX597. Waters depicted in green as stable and red as unstable

sites. The third stage is calculating the thermodynamic property of each water molecule. The calculation corresponds to the average of excess enthalpy, entropy, and free energy. Negative ΔH value of hydration site results in stronger interaction with the nearby water molecules and proteins in solution (e.g., charged group). Similarly, a positive ΔH value of hydration site represents weaker interactions with the adjacent water molecules and proteins in solution (e.g., near a hydrophobic residue). The hydration site (number) will give you the extent of volume measure. Meanwhile, hydration sites energies result in the enclosure and hydrophobic/hydrophilic balance of binding sites (Fig. 8.2). The net water transfer energy from the binding site to bulk, and from bulk to the binding site, contributes a broad ligand binding free energy. We majorly focus on a tiny (drug-sized) cluster of binding sites with highly unstable hydration sites. There are four following choices for designing a ligand:

(a) Displace: Hydration site has both ΔG and $\Delta H >> 0$ kcal mol^{-1}, generally favorable to displace (hydrophobic regions).
(b) Replace: Hydration sites with $\Delta H << 0$ kcal mol^{-1} but with $\Delta G \gtrsim 0$ kcal mol^{-1} are candidates for replacement.
(c) Interact: Highly conserved hydration sites, forming bridging waters.
(d) Avoid: When $\Delta G << 0$ kcal mol^{-1}, water molecules are highly stable, such cases it could be easier to avoid.

8.3.11 Quantitative WaterMap

WaterMap scoring considers only the ΔG. The contribution of each hydration site from the WaterMap is summarized using the simple equation.

$$\Delta G \Sigma_{i \epsilon \text{Atoms}} \Sigma_{s \epsilon \text{Sites}} O(a, s) \cdot \Delta G_s \tag{8.1}$$

where $O(a,s)$ measures the overlap of the atom "a" with site "s." The water scoring equation was taken from WaterMap Schrödinger (Garcia-Sosa and Mancera 2003).

WM/MM scoring calculation is similar to MM-GBSA calculation implemented in Prime (Prime, Schrödinger, LLC, New York, NY, 2019), which uses the advanced VSGB solvation models and optimized potential for liquid simulations (OPLS) 2.1 force field. For WaterMap simulations, the binding site was defined by the position of a co-crystal ligand in the protein and ligand site was used to describe the volume for the calculation of hydration sites that was carried out by thermodynamic analysis (Biswal et al. 2020). The Holo-WaterMap Scoring is relatively simple; integrating the continuous WaterMap within the vicinity of the ligand gives an estimate of the ligand effect on the solvent. Druggability, activity, and selectivity analyses will all be used for the ligand-binding site. WaterMap is used for more quantitative ligand scoring and can be calculated using the equation mentioned above (Eq. 8.1).

8.3.12 Druggability and Selectivity Using WaterMap

Druggability is a term used in drug discovery where the biological target binds with a drug molecule (binding pocket) (Kwon 2012). The concept of druggability is restricted to a small molecule, but it also extended to biomedical products, e.g., monoclonal antibodies (Stockwell and Roark 2011). It is estimated that 10–15% of human proteins are disease linked and druggable, 1–25% of disease-modifying are likely to be druggable. There are several tools available for predicting the nature of druggability (Stockwell 2011).

There is no single approach that will deliver a good target on demand; potential starting points can emerge from methods, namely genome sequencing studies, genomic screens, phenotypic screening, and existing drugs (Lansdowne 2018). WaterMap tool can be used to find the druggability of the compound, and the hydration site thermodynamics of a binding site also provides similar information. Protein structures, which are constant and high energy, are distributed on hydration sites besides being helpful in the identification of ligand-binding sites. Druggable cavities have unstable water molecules, and hydration sites provide the measure of the volume, whereas the energies of hydration site evidence the hydrophobic/hydrophilic balance of a binding site. A binding site with drug size small cluster comprises highly unstable hydration sites (Beuming et al. 2012). Usually, the drug-sized molecule contains the significant binding from occupying the binding site. The binding sites with unstable hydration sites are likely to be shallow or polar to bind a drug-like molecule. The binding site contains many stable water molecules and mostly hydrophobic in character. There are several methods, namely WaterMap, SZMAP, and WaterFLAP, for druggability prediction. Kohlmann et al. have applied MM-GB/SA and WaterMap methods for affinity calculation.

8.3.13 DOWSER^{++}

DOWSER uses a semi-empirical modification program for protein hydration based on average energy from MD simulation (Morozenko and Stuchebrukhov 2016). The position of the water molecules will be predicted and compared with the experimental data. A cutoff of -10 kcal mol^{-1} is used to determine how much water is occupied in the interior cavities before MD simulations. The script WaterDock helps to discover the number of potential water locations, which we further analyze with the methods implemented in the Dowser code previous version. Thus, the main reason for using WaterDock is to locate internal cavities that are unnoticed by the Dowser program. With the help of the PDB file, internal cavities present in the protein with a water probe of 1 Å and water molecules will be filled based on the minimized energies. Specific model (water-protein interaction) will be used for evaluating the water energies. This method will predict the water molecules present in the X-ray structure (Carugo and Bordo 1999; Zhang and Hermans 1996; Morozenko et al. 2014).

8.3.14 WaterDock

WaterDock is an algorithm that uses the freely available AutoDock Vina tool (Trott and Olson 2010) to predict the location of ordered water molecules in ligand-binding sites to very high accuracy (Sridhar et al. 2017). WaterDock was confirmed against high-resolution crystal structures, neutron diffraction data, and MD simulations approach. For justification, a set of proteins (14 structures of OppA structure) with high-resolution X-ray structures were used, and it predicts 88% of "consensus" water sites with a mean error of 0.78 Å using Acqua Alta method (Rossato et al. 2011). WaterDock is accurate and predicts 97% of the ordered water molecules, with a standard of 1 false-positive per structure. Ligand functional groups will be identified first, and their hydration sites were done based on semi-empirical function. "Favorable" protein–water interactions can also be identified. Two probabilistic water molecules are developed to predict WaterDock predictions by using data mining, heuristic, and machine learning techniques. The individual hydration of functional groups was first calculated from MD simulations of ligands (Ross et al. 2012).

8.3.15 WATsite: Hydration Site Prediction Program with PyMOL Interface

A graphical user interface (GUI) developed with inbuilt PyMOL plug-in free of charge for calculating the thermodynamic properties and hydration sites with the help of enthalpy and entropy of the water molecule. MD simulation can be performed, followed by hydration site identification and free energy estimations. This tool will solvate the proteins with explicit water molecules and performs MD simulations, where the fluctuations of water molecules in the protein-binding site

will be analyzed. The clustering algorithm and quality threshold (QT) (Glenn 2001) were applied for the prediction of locations of water molecules, which is followed by hydration sites that are used for constructing the pharmacophore models and enrichment analysis. Therefore, if a ligand restores the water present in the hydration site, it can be expected to increase the binding free energy. The "occupancy" term denotes the probability of a water molecule was observed in the hydration site using MD simulations approach. This tool will help in the predicted hydration sites and estimate the desolvation free energies occupied by the replacing water molecules present in the protein–ligand binding site within the user-specified distance (default value 1 Å). The main aim of this tool is to estimate the desolvation free energy of water molecules present in the protein-binding site. Once the ligand site is fixed, the water molecules are retained for simulation. The energy minimization was performed using the steepest descent algorithm with the periodic boundary conditions. The binding free energy for protein includes other important contributions, such as the direct protein–ligand interaction energy and desolvation energy of the ligand.

The desolvation free energy of each hydration site is determined by analyzing the enthalpy and entropy of the water molecules inside a hydration site.

$$\Delta G_{\mathrm{hs}} = \Delta H_{\mathrm{hs}} - T\Delta S_{\mathrm{hs}} \tag{8.2}$$

where ΔH_{hs} and ΔS_{hs} denote the enthalpic and entropic changes of water molecules transferred from the bulk solvent into the hydration site of the protein cavity.

$$\Delta H_{\mathrm{hs}} \approx \Delta E_{\mathrm{hs}} = E_{\mathrm{hs}} - E_{\mathrm{bulk}} \tag{8.3}$$

where E_{hs} term denotes the interaction energy of a water molecule in the hydration site, based on the average sum of van der Waals and electrostatic interactions between each water molecule inside a given hydration site with the protein and all the other water molecules. E_{bulk} is the interaction energy of a water molecule with its surrounding environment in the bulk solvent.

Assuming no alteration at the moment apart of the partition function on transferring a water molecule from the bulk solvent into the protein cavity, ΔS_{hs} can be estimated by

$$\Delta S_{\mathrm{hs}} = R \ln \left({}^{C^{\circ}}\!/_{8\pi^2} \right) - R \int \mathrm{pext}(q) \ln\ \mathrm{pext}(q) dq \tag{8.4}$$

where C° is the concentration of pure water (1 molecule/29.9 Å^3), R is the gas constant, and pext (q) is the external mode probability density function of the water molecules' translational and rotational motions during the MD simulation (Hu and Lill 2014).

8.4 Selectivity

There is an essential role in the binding site of hydration in the ligand selectivity, as analyzed for kinase target in the earlier study (Robinson et al. 2010). Displacement of potential water molecules site has a more significant impact on affinity and selectivity. Longer chains in kinases determine the thermodynamically unstable hydration sites (Knegtel and Robinson 2011), resulting in binding affinity gain, and a slight change in the mobility of the hydration site towards the ligand corresponds to the selectivity of the kinases. Energetics and hydration sites will be useful to analyze the selectivity differences.

8.5 Limitations

The hydration site analysis is carried out based on the protein rigidness. Restraints will be challenging for hydration sites, where the number and strength of hydrogen bonds are affected by the variability of hydrogen-bonding partners. Similarly, inappropriate binding site analysis will cause flexibility in proteins. Restricted binding sites might have difficulties to be solvated as the buried sites cannot be reached without conformational rearrangements. WaterMap will compare all the energies for each hydration sites using thermodynamic approaches. It does not provide absolute binding free energies, but offer only relative free energy changes upon water displacement. The estimation for relative contributions of individual hydration sites to the macroscopic thermodynamic property is formulated as a sum of water enthalpy and entropy from MD using different theoretical approaches. WaterMap provides a comparative ranking of similar ligands and finds the linking between the binding site desolvation and affinity. Any interpretation of the structure–activity relationship (SAR) is only meaningful if the binding is primarily due to hydrophobic and not electrostatic interactions. Therefore, WaterMap is not a replacement for scoring functions or advanced binding affinity estimations.

8.6 Conclusions

Water molecules are essential for protein–ligand interactions and the study of the thermodynamic properties during binding. The chapter covered the significance of hydration sites in conferring stability, influencing protein folding, and key biomolecular interactions. Vital intramolecular forces such as hydrophobic, hydrophilic, electrostatic, and hydrogen bond network aid to establish the interaction of the first-order hydrating water molecules to bulk and the protein environment. Likewise, thermodynamic signatures such as the entropy and enthalpic contributions about the mobility of the water and the influence sought by the environment are contributing factors for water displacements. Computational approaches like prediction tools have been discussed here. In the recent era, hydration site analysis is an essential criterion for structure-based drug design projects for further exploration of the

binding site profiling will highlight the fundamental interactions. This chapter depicts the importance of hydration sites in the SAR or analogs around the lead compound might help in designing lead molecules and optimization. In structure-based drug design, potency and ADMET properties are important factors to water interactions; this helps in understanding the molecular detection of protein–ligand complexes. There is a significant importance using computational tools to realize a lot about water role and signifying opportunities in structure-based design. We summarized concepts and critical applications towards hydration sites, which are useful in drug design projects.

Acknowledgement JJ thank the DST INDO-TAIWAN (GITA/DST/TWN/P-86/2019 dated: 04/03/2020), Board of Research in Nuclear Sciences (BRNS) (35/14/ 02/2018 BRNS/35009), Indian Council for Medical Research (ICMR) (No.BIC/12(07)/2015), DST-Science and Engineering Research Board (SERB) (No.EMR/2016/000498), UGC Research Award (No. F. 30-32/2016 (SA-II) dt 18.04.2016), DST- Fund for Improvement of S&T Infrastructure in Universities & Higher Educational Institutions (FIST) (SR/FST/LSI-667/2016) (C), DST-Promotion of University Research and Scientific Excellence (PURSE) (No. SR/PURSE Phase 2/38 (G), 2017 and MHRD-RUSA 2.0, New Delhi (F.24-51/2014-U, Policy (TNMulti-Gen), Dept. of Edn. Govt. of India, Dt.09.10.2018). JB is grateful to the UGC OBC National Fellowship (F./2015-16/NFO-2015-17-OBC-PON-29027)) A special word of gratitude to Mr. Raghu Rangasamy, Vice-President, and the Application Scientists of Schrodinger, India, for their timely help and suggestions.

References

Abel R, Young T, Farid R, Berne BJ, Friesner RA (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. J Am Chem Soc 130:2817–2831

Ahlrichs R, Bar M, Haser M, Horn H, Kolmel C (1989) Electronic structure calculations on workstation computers: the program system turbomole. Chem Phys Lett 162(3):165–169

Anandakrishnan R, Izadi S, Onufriev AV (2019) Why computed protein folding landscapes are sensitive to the water model. J Chem Theory Comput 15(1):625–636

Ansari A, Jones CM, Henry ER, Hofrichter J, Eaton WA (1992) The role of solvent viscosity in the dynamics of protein conformational changes. Science 256(5065):1796–1798

Baldwin RL (2014) Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. Proc Natl Acad Sci U S A 111(36):13052–13056

Ball P (2008) Water as an active constituent in cell biology. Chem Rev 108(1):74–108

Bandyopadhyay S, Chakraborty S, Bagchi B (2005a) Secondary structure sensitivity of hydrogen bond lifetime dynamics in the protein hydration layer. J Am Chem Soc 127(47):16660–16667

Bandyopadhyay S, Chakraborty S, Balasubramanian S, Bagchi B (2005b) Sensitivity of polar solvation dynamics to the secondary structures of aqueous proteins and the role of surface exposure of the probe. J Am Chem Soc 127(11):4071–4075

Baroni M, Cruciani G, Sciabola S, Perruccio F, Mason JS (2007) A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): theory and application. J Chem Inf Model 47(2):279–294

Bellissent-Funel MC, Hassanali A, Havenith M, Henchman R, Pohl P, Sterpone F, van der Spoel D, Xu Y, Garcia AE (2016) Water determines the structure and dynamics of proteins. Chem Rev 116(13):7673–7697

Beuming T, Che Y, Abel R, Kim B, Shanmugasundaram V, Sherman W (2012) Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. Proteins 80(3):871–883

Biswal J, Jayaprakash P, Suresh Kumar R, Venkatraman G, Poopandi S, Rangasamy R, Jeyaraman J (2020) Identification of Pak1 inhibitors using water thermodynamic analysis. J Biomol Struct Dyn 38(1):13–31

Bizzarri AR, Cannistraro S (2002) Molecular dynamics of water at the protein-solvent interface. J Phys Chem B 106(26):6617–6633

Bodnarchuk MS (2016) Water, water, everywhere... It's time to stop and think. Drug Discov Today 7:1139–1146

Boyarkin OV, Kosheleva MA, Aseev O, Maksyutenko P, Rizzo TR, Zobov NF, Lodi L, Tennyson T, Polyansky OL (2013) Accurate bond dissociation energy of water determined by triple-resonance vibrational spectroscopy and *ab initio* calculation. Chem Phys Lett 568–569:14–20

Cao Z, Bowie JU (2014) An energetic scale for equilibrium H/D fractionation factors illuminates hydrogen bond free energies in proteins. Protein Sci 23(5):566–575

Carugo O, Bordo D (1999) How many water molecules can be detected by protein crystallography? Acta Crystallogr D Biol Crystallogr 55(Pt 2):479–483

Černy J, Schneider B, Biedermannová L (2017) WatAA: atlas of protein hydration. Exploring synergies between data mining and ab initio calculations. Phys Chem Chem Phys 19 (26):17094–17102

Chakraborty S, Sinha SK, Bandyopadhyay S (2007) Low-frequency vibrational spectrum of water in the hydration layer of a protein: a molecular dynamics simulation study. J Phys Chem B 111 (48):13626–13631

Chaplin MF (2008) Water in biological recognition processes. In: Begley TP (ed) Wiley Encyclopedia of chemical biology. Wiley, New York, pp 1–8

Cheung MS, Garcia AE, Onuchic JN (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc Natl Acad Sci U S A 99(2):685–690

Colin RG, Lan JB, Matthew PL, Suzanna CW (2016) The Cambridge structural database. Acta Crystallor B Struct Sci Cryst Eng Mater 72(2):171–179

Covalt JC Jr, Roy M, Jennings PA (2001) Core and surface mutations affect folding kinetics, stability and cooperativity in IL-1 beta: does alteration in buried water play a role? J Mol Biol 307(2):657–669

De Beer SB, Vermeulen NP, Oostenbrink C (2010) The role of water molecules in computational drug design. Curr Top Med Chem 1:55–66

Fenimore PW, Frauenfelder H, McMahon BH, Parak FG (2002) Slaving: solvent fluctuations dominate protein dynamics and functions. Proc Natl Acad Sci U S A 99(25):16047–16051

Fernández A, Kardos J, Goto Y (2003) Protein folding: could hydrophobic collapse be coupled with hydrogen-bond formation? FEBS Lett 536(1–3):187–192

Finkelstein A, Ptitysyn O (2016) Protein physics: a course of lectures, 2nd edn. Elsevier, Academic

Fusani L, Wall I, Palmer D, Cortes A (2018) Optimal water networks in protein cavities with GAsol and 3D-RISM. Bioinformatics 34(11):1947–1948

Garcia-Sosa AT, Mancera RL, Dean PM (2003) WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. J Mol Model 9(3):172–182

Glenn F (2001) A comprehensive overview of basic clustering algorithms. Comp Sci 2:1–37

Glonek GFV, McCullagh P (1995) Multivariate logistic models. J R Stat Soc 57(3):533–546

Grant JA, Pickup BT, Nicholls A (2001) A smooth permittivity function for poisson boltzmann solvation methods. J Comput Chem 22(6):608–640

Gromiha MM (2000) Protein bioinformatics: from sequence to function, 1st edn. Elsevier, Academic

Gummadi SN (2003) What is the role of thermodynamics on protein stability? Biotechnol Bioprocess Eng 8:9–18

Hildebrandt A, Blossey R, Rjasanow S, Kohlbacher O, Lenhof HP (2007) Electrostatic potentials of proteins in water: a structured continuum approach. Bioinformatics 23(2):e99–e103

Hu B, Lill MA (2014) WATsite: hydration site prediction program with PyMOL interface. J Comput Chem 35(16):1255–1260
Huggins DJ (2016) Studying the role of cooperative hydration in stabilizing folded protein states. J Struct Biol 196(3):394–406
Ikura T, Urakubo Y, Ito N (2004) Water-mediated interaction at a protein- protein interface. Chem Phys 307(2–3):111–119
Jana M, Bandyopadhyay S (2012) Conformational flexibility of a protein-carbohydrate complex and the structure and ordering of surrounding water. Phys Chem Chem Phys 14(18):6628–6638
Jukic M, Konc J, Gobec S, Janezic D (2017) Identification of conserved water sites in protein structures for drug design. J Chem Inf Model 57(12):3094–3103
Jurecka P, Cerny J, Hobza P, Salahub DR (2007) Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. J Comput Chem 28(2):555–569
Kangueane P, Nilofer C (2018) Protein-protein and domain-domain interactions, 1st edn. Springer, New York
Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652
Knegtel RM, Robinson DD (2011) A role for hydration in interleukin-2 inducible T cell kinase (Itk) selectivity. Mol Inform 30(11–12):950–959
Kubinyi H (2001) Hydrogen bonding: the last mystery in drug design. Pharmacokinetic optimization in drug research, 1st edn. Wiley, New York
Kuffel A (2017) How water mediates the long-range interactions between remote protein molecules. Phys Chem Chem Phys 19(7):5441–5448
Kwon B (2012) Chemical biologist targets 'undruggable' proteins linked to cancer in quest for new cures. Brent Stockwell interview Medical Xpress. Accessed 17 May 2012
Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9(1):23–27
Lagasse HA, Alexaki A, Simhadri VL, Katagiri NH, Jankowski W, Sauna ZE, Kimchi-Sarfaty C (2017) Recent advances in (therapeutic protein) drug development. F1000Res 1:6–113
Lansdowne LE (2018). Target identification & validation in drug discovery. https://www.technologynetworks.com/drug-discovery/articles/target-identification-validation-in-drug-discovery-312290. Accessed 20 Mar 2018
Lazaridis T (1998) Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory Chemistry 102(18):3531–3541
Lehmkuhler F, Forov Y, Elbers M, Steinke I, Sahle CJ, Weis C, Tsuji N, Itou M, Sakurai Y, Poulain A, Sternemann C (2017) Temperature dependence of the hydrogen bond network in trimethylamine N-oxide and guanidine hydrochloride-water solutions. Phys Chem Chem Phys 19(41):28470–28475
Levinson NM, Boxer SG (2014) A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. Nat Chem Biol 10(2):127–132
Levy Y, Onuchic JN (2004) Water and proteins: a love-hate relationship. Proc Natl Acad Sci U S A 101(10):3325–3326
Lim VI, Curran JF, Garber MB (2012) Hydration shells of molecules in molecular association: a mechanism for biomolecular recognition. J Theor Biol 301:42–48
Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285(5):2177–2198
Lopez ED, Arcon JP, Gauto DF, Petruk AA, Modenutti CP, Dumas VG, Marti MA, Turjanski AG (2015) WATCLUST: a tool for improving the design of drugs based on protein-water interactions. Bioinformatics 31(22):3697–3699
Lu Y, Wang R, Yang CY, Wang S (2007) Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. J Chem Inf Model 47(2):668–675
Malleshappa Gowder S, Chatterjee J, Chaudhuri T, Paul K (2014) Prediction and analysis of surface hydrophobic residues in tertiary structure of proteins. Sci World J 2014:971258

Manzoni F, Ryde U (2018) Assessing the stability of free-energy perturbation calculations by performing variations in the method. J Comput Aided Mol Des 32(4):529–536
Maurer M, Oostenbrink C (2019) Water in protein hydration and ligand recognition. J Mol Recognit 32(12):e2810
Mondal J, Friesner RA, Berne BJ (2014) Role of desolvation in thermodynamics and kinetics of ligand binding to a kinase. J Chem Theory Comput 10(12):5696–5705
Morozenko A, Stuchebrukhov AA (2016) Dowser++, a new method of hydrating protein structures. Proteins 84(10):1347–1357
Morozenko A, Leontyev IV, Stuchebrukhov AA (2014) Dipole moment and binding energy of water in proteins from crystallographic analysis. J Chem Theory Comput 10(10):4618–4623
Nakasako M (2004) Water-protein interactions from high-resolution protein crystallography. Philos R Soc Lond B: Biol Sci 359(1448):1191–1204
Novak U, Grdadolnik J (2017) The hydration of concanavalin a studied by infrared spectroscopy. J Mol Struct 1135:138–143
Oleinikova A, Smolin N, Brovchenko I, Geiger A, Winter R (2005) Formation of spanning water networks on protein surfaces via 2D percolation transition. J Phys Chem B 109(5):1988–1998
Olsson TS, Williams MA, Pitt WR, Ladbury JE (2008) The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. J Mol Biol 384(4):1002–1017
Panigrahi SK, Desiraju GR (2007) Strong and weak hydrogen bonds in the protein-ligand interface. Proteins 67(1):128–141
Pappas CG, Tzakos AG, Gerothanassis IP (2012) On the hydration state of amino acids and their derivatives at different ionization states: a comparative multinuclear NMR and crystallographic investigation. J Amino Acids 2012:565404
Parsegian VA (2002) Protein-water interactions. Int Rev Cytol 215:1–31
Perutz MF, Kendrew JC, Watson HC (1965) Structure and function of haemoglobin: II some relations between polypeptide chain configuration and amino acid sequence. J Mol Biol 13 (3):669–678
Pradhan MR, Nguyen MN, Kannan S, Fox SJ, Kwoh CK, Lane DP, Verma CS (2019) Characterization of hydration properties in structural ensembles of biomolecules. J Chem Inf Model 59 (7):3316–3329
Ragone R (2001) Hydrogen-bonding classes in proteins and their contribution to the unfolding reaction. Protein Sci 10:2075–2082
Rashin AA, Bukatin MA (1991) Continuum based calculations of hydration entropies and the hydrophobic effect. J Comput Chem 95(8):2942–2944
Raymer ML, Sanschagrin PC, Punch WF, Venkataraman S, Goodman ED, Kuhn LA (1997) Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. J Mol Biol 265(4):445–464
Reddy CK, Das A, Jayaram B (2001) Do water molecules mediate protein-DNA recognition? J Mol Biol 314(3):619–632
Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS (2004) Simulations of the role of water in the protein-folding mechanism. Proc Natl Acad Sci U S A 101(17):6456–6461
Robinson DD, Sherman W, Farid R (2010) Understanding kinase selectivity through energetic analysis of binding site waters. ChemMedChem 5(4):618–627
Rodier F, Bahadur RP, Chakrabarti P, Janin J (2005) Hydration of protein-protein interfaces. Proteins 60(1):36–45
Rose GD, Wolfenden R (1993) Hydrogen bonding, hydrophobicity, packing, and protein folding. Annu Rev Biophys Biomol Struct 22:381–415
Ross GA, Morris GM, Biggin PC (2012) Rapid and accurate prediction and scoring of water molecules in protein binding sites. PLoS One 7(3):e32036
Rossato G, Ernst B, Vedani A, Smiesko M (2011) Acqua Alta: a directional approach to the solvation of ligand-protein complexes. J Chem Inf Model 51(8):1867–1881
Ruhmann E, Betz M, Fricke M, Heine A, Schafer M, Klebe G (2015) Thermodynamic signatures of fragment binding: validation of direct versus displacement ITC titrations. Biochim Biophys Acta 1850(4):647–656
Santa Fe NM (2013) SZMAP 1.5.0.2: OpenEye Scientific Software

Schneider B, Cohen DM, Schleifer L, Srinivasan AR, Olson WK, Berman HM (1993) A systematic method for studying the spatial distribution of water molecules around nucleic acid bases. Biophys J 65(6):2291–2303
Schrödinger Release (2019) Prime, Schrödinger. LLC, New York, NY
Schrödinger Release (2020) 2020–1: WaterMap, Schrödinger. LLC, New York, NY
Seyedi S, Matyushov D (2018) Dipolar susceptibility of protein hydration shells. Chem Phys Lett 713:210–214
Shanthi V, Rajesh CK, Jayalakshmi J, Vijay VG, Sekar K (2002) WAP: water analysis package—a web-based package to calculate geometrical parameters between water oxygen and protein atoms. J Appl Crystallogr 36(1):167–168
Smolin N, Oleinikova A, Brovchenko I, Geiger A, Winter R (2005) Properties of spanning water networks at protein surfaces. J Phys Chem B 109(21):10995–11005
Snyder PW, Lockett MR, Moustakas DT, Whitesides GM (2014) Is it the shape of the cavity, or the shape of the water in the cavity? Eur Phys J Spec Top 223:853–891
Spyrakis F, Ahmed MH, Bayden AS, Cozzini P, Mozzarelli A, Kellogg GE (2017) The roles of water in the protein matrix: a largely untapped resource for drug discovery. J Med Chem 60 (16):6781–6827
Sridhar A, Ross GA, Biggin PC (2017) Waterdock 2.0: water placement prediction for Holo-structures with a pymol plugin. PLoS One 12(2):e0172743
Stockwell B (2011) Outsmarting Cancer. A biologist talks about what makes disease-causing proteins so difficult to target with drugs. Sci Am 305(4):20
Stockwell B, Roark B (2011) The quest for the cure: the science and stories behind the next generation of medicines. Columbia University Press, New York
Takano K, Yamagata Y, Yutani K (2003) Buried water molecules contribute to the conformational stability of a protein. Protein Eng 16(1):5–9
Tanger JC, Pitzer KS (1989) Calculation of the thermodynamic properties of aqueous electrolytes to 1000.degree.C and 5000 bar from a semi continuum model for ion hydration. J Phys Chem 93 (12):4941–4951
Tarcsay Á, Keserű GM (2015) Is there a link between selectivity and binding thermodynamics profiles? Drug Discov Today 1:86–94
Thomas L (2020) "Protein Structure Determination". News-Medical. https://www.news-medical.net/life-sciences/Protein-StructureDetermination.aspx. Accessed 20 Mar 2020
Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91
Trott O, Olson AJ (2010) AutoDockVina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461
Van Dijk E, Hoogeveen A, Abeln S (2015) The hydrophobic temperature dependence of amino acids directly calculated from protein structures. PLoS Comput Biol 11(5):e1004277
Virtanen JJ, Sosnick TR, Freed KF (2014) Ionic strength independence of charge distributions in solvation of biomolecules. J Chem Phys 141(22):22D503
Wüthrich K, Otting G, Liepinsh E (1992) Protein hydration in aqueous solution. Faraday Discuss 93:35–45
Yokomizo T, Nakasako M, Yamazaki T, Shindo H, Higo J (2005) Hydrogen-bond patterns in the hydration structure of a protein. Chem Phys Lett 401(4–6):332–336
Zapadka KL, Becher FJ, Gomes Dos Santos AL, Jackson SE (2017) Factors affecting the physical stability (aggregation) of peptide therapeutics. Interf Focus 7(6):20170030
Zhang L, Hermans J (1996) Hydrophilicity of cavities in proteins. Proteins 24(4):433–438
Zhong D, Pal SK, Zewali AH (2011) Biological water: a critique. Chem Phys Lett 503(1–3):1–11

Molecular Dynamics Simulation: Methods and Application

9

Sakshi Singh and Vinay Kumar Singh

Abstract

The complexity of the 3D structure of a protein is still challenging in the area of structural biology. Thermodynamics-based methods, including molecular dynamics (MD) simulations, enable our understanding of protein's conformational detail at the atomic level. Proteins are flexible molecules. MD simulation provides information about the dynamic perturbations that occur in a protein or protein–ligand complex. As compared to docking, MD simulation also considers the physiological conditions such as temperature, pH, presence of water, ions, and other molecules of the system. The force field and the software packages are several choices involved in large-scale simulation, which analyze the single protein, protein–ligand, and protein–protein structure, respectively. In this chapter, we discuss all the approaches along with packages that cover the entire molecular-level to cellular-level features of proteins. MD simulation approaches are very useful in assessing the stability of a protein model or protein–ligand complex and mutational study as well.

Keywords

Protein · Force field · Molecular dynamics · Energy minimization · Root mean square deviation · Root mean square fluctuation

S. Singh
Department of Molecular and Human Genetics, Institute of Science, Banaras Hindu University, Varanasi, India

V. K. Singh (✉)
Centre for Bioinformatics, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_9

9.1 Introduction

Molecular modeling defines the characteristics of a three-dimensional (3D) structure of a protein and the structure–function relationship by the use of an *in silico* approach/computer-based approach (Pensak 1989; Forster 2002). 3D structures of proteins and their complexes provide valuable information to understand the mechanism of the machinery of life and also screening and designing new drugs in therapeutic aspects. Some methods applied in molecular modeling are automatic structure generation, analysis of 3D databases, homology modeling, *ab initio* modeling, diversity analysis, molecular docking, mutational analysis, etc. (Breda et al. 2007; Kalita et al. 2019a, b). Computational chemistry and their methodologies, including energy minimization, free-energy calculation of a molecular system (Shukla et al. 2019), and Monte Carlo (MC) simulation, are the nucleus for molecular modeling (Kantarci-Carsibasi et al. 2008). The MC simulation algorithms can run on any computer, which creates a large number of random numbers from a selected distribution. First, the elements are represented, and the corresponding distribution is chosen for the simulation. Then, the mathematical calculations for these elements are determined, and the number (N) of simulation runs have to be determined (Gentle 2009). The simulation produces N random numbers that follow a given distribution of each of the elements and computes N times. To summarize the results of these calculations, average, variance, and confidence intervals can be calculated. MC simulation is a method that allows obtaining results when calculating the mathematical problem and/or it is challenging to find an analytical solution that is too complex (Hubbard and Samuelson 2009).

The other important simulation approach is molecular dynamics (MD) simulation, which studies the temporal evolution of the coordinates and the state of the structure of a given macromolecule (Kalita et al. 2019a, b; Paquet and Viktor 2015; Sonkar et al. 2019). This evolution is called a trajectory. The trajectory obtained by solving Newton's laws of motion is important to estimate numerous time-dependent observatories, such as targeted molecular surface, the interaction between a small molecule and a glycoprotein, the epitope–paratope interaction between the antigen and the antibody, the presence and disappearance of a particular channel or cavity, and the fusion of the glycoprotein membrane (Paquet and Viktor 2015; Salmaso and Moro 2018). The study of protein and other polymers, MD simulation, was primarily introduced to leverage the increasing computing ability in the early 1980s (Adcock and McCammon 2006). It was used with a large number of different estimations while studying the protein-folding problem, such as the impact of protein dynamics on catalysis and ligand binding (Salsbury 2010).

Nearly four decades since the first MD simulation appeared, we have learned a considerable amount about how different signaling pathways within the human system can be affected by the protein structure and their function. Disrupting a portion or complete 3D protein structure can cause apparent aberrations in cellular disease-causing processes or activity loss in the case of an enzyme (Kalita et al. 2018; Shukla et al. 2018a, b, c, d, e), thus illustrating the need to predict the 3D protein structure and recognize its biophysical properties. MD simulation was first

Table 9.1 The important landmarks of the development of MD simulation

Scientist	Milestones/achievements	Years
Metropolis	The first introduction of MC simulation of liquids (hard spheres)	1953
Wood	First MC simulation with Lennard-Jones potential	1957
Alder	First MD simulation of a liquid (hard spheres)	1957
Rahman	First MD simulation with Lennard-Jones potential	1964
Karplus and McCammon	First MD simulation of proteins	1977
Karplus	The CHARMM general-purpose force field and the MD program	1983
Kollman	The AMBER general-purpose force field and MD program	1984
Car-Parrinello	First full quantum mechanics (QM) simulations	1985
Kollmann	First quantum mechanics/molecular mechanics (QM-MM) simulation	1986

used in the late 1950s by Alder and Wainwright to analyze the hard-sphere system (Alder and Wainwright 1959, 1960). The next major breakthrough came in 1964 when Rahman used the first simulation with a working functional liquid argon potential (Rahman 1964). In the late 1970s, first MD simulation of protein was carried out and later its significance was awarded by the 2013 Nobel Prize in Chemistry (McCammon et al. 1977; Levitt and Lifson 1969). The brief description of simulation milestones is depicted in Table 9.1.

A theoretical method describes the MD simulation, which provides information about the time-dependent behavior of the molecular target system by combining Newton's laws of motion. The structure at the atomic level provides insight into how the protein or other biomolecule functions. Nevertheless, the MD simulation approach based on a general concept of interatomic interactions defines how atoms move over time in a protein or protein complex system (Karplus and McCammon 2002). MD simulation approach can observe the biomolecular processes, viz. conformation changes, protein–ligand binding, and protein–peptide binding showing the orientation of all atoms at femtosecond (fs) spatial resolution (Dror et al. 2010; Hollingsworth and Dror 2018). The simulations can also reveal the involvement of certain alterations (such as mutation, phosphorylation, protonation, or the addition or removal of a ligand) depending on protein or protein-ligand complex.

The mechanism of MD simulation is capable of identifying interactions at the atomic level between all the system peripherals, acting as a "computational microscope" (Lee et al. 2009; Ingólfsson et al. 2016). It included numerically solved classical equations of motion in its most basic form over a given period of time. The resulting time series, called trajectory, that afterward be visualized and analyzed in detail. MD simulations have gotten increasingly useful for the experimental molecular biologists' point of view as of late. In the theoretical structural biology, simulations have started to appear regularly, where they used both to depict experimental findings and to direct experimental research. The methods underlying MD

simulations have evolved evermore and show the future application in the simulation field.

9.2 The Theory Behind MD Simulation

MD simulation was designed to study the time evolution of protein conformations and also provide the kinetic and thermodynamic information with detailing of motion of each atom as a function of time (Adcock and McCammon 2006; Hollingsworth and Dror 2018). Classical MD simulation depicts the movements of the considerable number of particles in a molecular system by comprehending Newton's equations of motion (Hug 2013).

The second law of Newton's equation of motion is given for each atom i,

$$F_i = m_i a_i \tag{9.1}$$

where F is the net force applied on the particle, m is the mass of the particle, and a is the acceleration of the particles, respectively.

$$F_i(t) = m_i \frac{d^{(2)} r_i(t)}{dt^2} \tag{9.2}$$

Here, $F_i(t)$ is at the time t the force on atom i, and it is also expressed as the gradient of the potential energy. r_i and m_i represent the position and mass of atom i, respectively.

Several numerical algorithms, such as the Verlet algorithm, Velocity Verlet algorithm, Leap-frog algorithm, and Beeman's algorithm, have been developed for integrating the equations of motion. MD simulation provides a wealth of quantitative information about the protein and peptide structures and their dynamics. However, the method has certain confinements, such as dependency on the size of the system and time frame, which is currently restricted to hundreds of nanoseconds or a few microseconds.

Proteins are typically represented using a model of an atomic level where almost all of the atoms are explicitly present (Hospital et al. 2015). MD simulation of protein and their complexes can be utilized for finding the answer to many types of questions. In this approach, several choices must be made, including the algorithm-based software and force field, where the protein molecules are involved in the simulation, and describe their behavior (Collier et al. 2020). It can provide the flexibility or mobility of various protein structure regions, the accuracy of the modeled structure, refine the 3D structure, and also define the protein biomolecular process (Fig. 9.1).

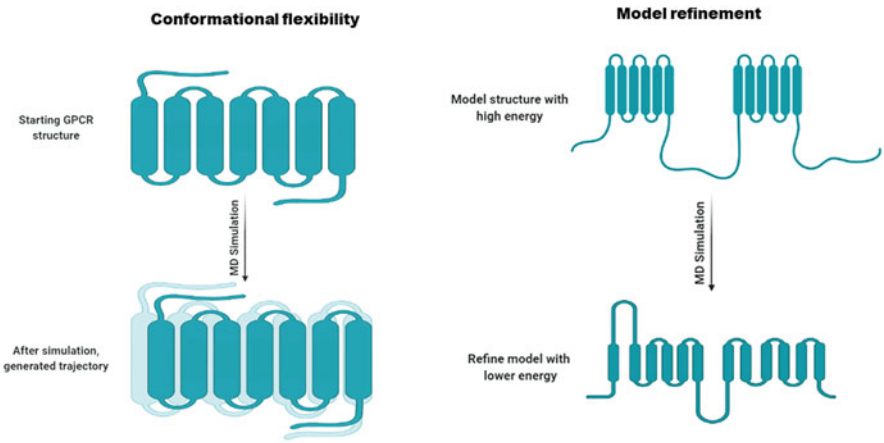


Fig. 9.1 Application of MD simulation. Studying conformational flexibility and refinement of the model structure

9.3 Methods of MD Simulation

As per the structure of protein molecules, i.e., the relative coordinates of the constituent atoms, it is possible to use different approaches to investigate the dynamics of that system (Beck and Daggett 2004; Hernández-Rodríguez et al. 2016). The study selection method is represented as a flowchart (Fig. 9.2). In this section, a number of those methods are defined and discussed.

9.3.1 Force Fields for MD Simulation

With the assistance of QM/MM, the potential energy function evaluates the protein conformation flexibility. Force field defines the set of potential energy functions from which the forces originated (Vanommeslaeghe and Guvench 2014). It indicates the functional form and parameter sets for MM and MD simulation to measure the potential energy of atoms or coarse-grained particles in a system (Guvench and MacKerell 2008; Lorenz and Doltsinis 2012). Computing the force field energy as a function of the 3D structure offers the full potential energy surface of a molecule for all possible 3D conformations (Smith et al. 2017). The force field contributions of the different atomic forces which govern MD simulation. Many force fields are widely used in the simulation, including AMBER, CHARMM, OPLS, and GROMOS. All are varied mainly in the manner in which they parameterize but generally give similar results.

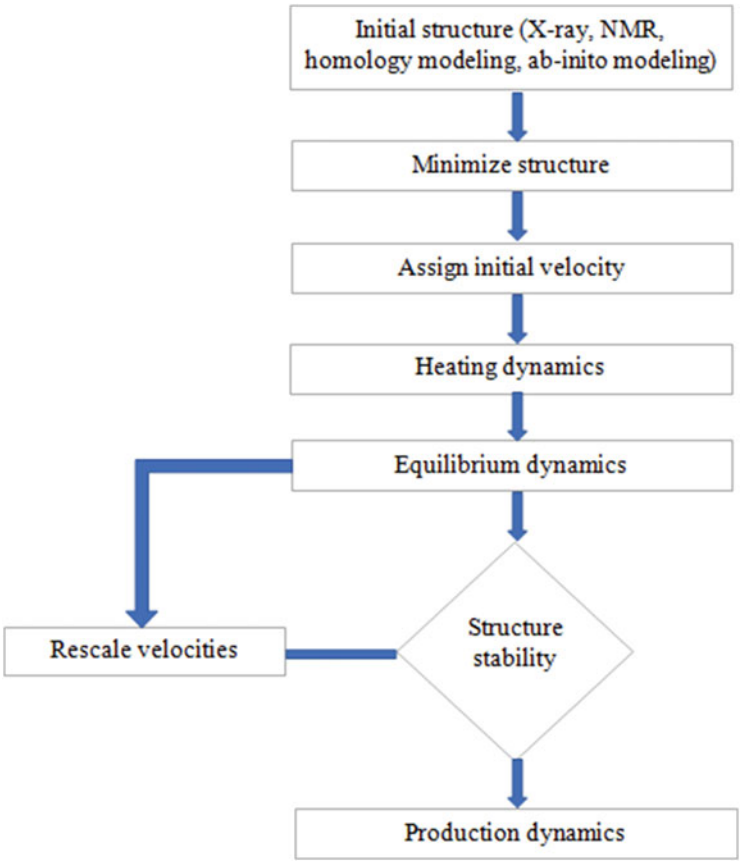


Fig. 9.2 A workflow diagram is representing the steps applied in MD simulation

9.3.1.1 CHARMM (Chemistry at Harvard Macromolecular Mechanics)

CHARMM is used for biomolecular simulations of protein, nucleic acid, and small molecules. Investigation of the structural details in the MD refinement strategy is inferred solvation parameters by weakly enforcing constraints on the application of secondary structures while enabling enough flexibility for rearrangement (Brooks et al. 2009; Zhu et al. 2012). CHARMM is one of the oldest systems in MD simulation. Free-energy disturbance (FEP), quasi-harmonic entropy estimation, correlation analysis, and hybrid quantum mechanics (QM/MM) methods are involved in this force field (Ryazantsev et al. 2019). The most widely used CHARMM force field such as CHARMM19 is used for the single atom, and CHARMM22 is used for all atoms (Vanommeslaeghe and MacKerell 2015). The force field of CHARMM27 has been ported to GROMACS and officially supported as from version 4.5.

9.3.1.2 OPLS (Optimized Potentials for Liquid Simulations)

OPLS is an ensemble of force fields developed by William L. Jorgensen (Macchiagodena et al. 2017). There are two types of force fields in OPLS: united atom (OPLS-UA) and all atoms (OPLS-AA). OPLS-AA force-field parameters were defined in the 1990s for nucleoside bases and polynucleotides (Sweere and Fraaije 2017). Parameters for torsional energetics were taken from small organic molecules, however, and the only published polynucleotide study was a duplex DNA dodecamer in simulations. OPLS force field evaluated receptor–ligand interactions for fungicides with minimum binding energy (Amir et al. 2018). OPLS-AA/M used as a basis for improved torsional energetics for proteins based on the density functional theory (DFT) calculations. Main features of OPLS are as follows:

1. Derived from AMBER (intramolecular)
2. Nonbonded terms optimized for small molecule solvation
3. No lone pair, no hydrogen bond term

9.3.1.3 AMBER (Assisted Model Building and Energy Refinement)

This force field is used to analyze the protein, nucleic acid, and carbohydrate biomolecules (Case et al. 2005). Three main steps such as system preparation, simulation, and trajectory analysis allow implementation in this force field. As of GROMACS version 4.5, the following seven AMBER force fields are natively supported: AMBER-94,96,99,99SB,99SB-ILDN,03, GS. Unlike the VMD graphics tools that can read our trajectory format for animation preparation, Amber tools can also be used for the NAMD simulation program and force fields to comprehend or elucidate the data in Amber's parameter topology files. Many of the characteristics of AMBER are "Vacuum" simulations, nonbonded cutoff simulations, and free-energy simulations for pairwise decomposable potentials (Case et al. 2005). A major benefit of the AMBER parameters for the simulation of complex membranes is the primitive state of the force field remaining. It applies a significant amount of development to make the force field easy to apply to a variety of lipids and lipid mixtures. Many developments in most popular force fields (CHARMM, AMBER, and OPLS-AA) have occurred (Geng et al. 2019). Main features of AMBER are as follows:

1. Few atoms available, some calculations are impossible;
2. Experimental data + quantum chemical calculations;
3. Possible explicit terms for hydrogen bonds and treatment of lone pairs.

9.3.1.4 GROMOS (GROningen Molecular Simulation)

This force field includes a careful parametrization of the unbonded interactions aimed at reproducing the thermodynamic properties of small molecules representing the most basic chemical functions and using them as building blocks to model more complex systems. The GROMOS range of parameters is based on the GROMOS 54A7 united-atom force field. Because of its ability to replicate condensed phase properties, the family of force fields was chosen for the parametrization strategy. Appropriately, the torsional mechanical molecular profiles were applied to quantum

mechanical derivatives. For the parametrization of the partial charges and checking against the thermodynamic properties of organic liquids, the molecular mechanical atomic charges were applied at the quantum mechanical dipole moments. GROMOS's methodology is to change the partial atomic charges for the reproduction of the thermodynamic properties. The force-field GROMOS 53A6 is ideal for describing the distribution of molecules within a large system. For example, GBP1 protein exhibits antiviral activity against the influenza virus and is involved in the interferon-gamma signaling pathway. GM-CSF cytokine stimulates the expression of GBP1 protein, and this docked complex system was an equilibrium state during implement the gromos53a6 force field (Singh et al. 2020).

9.3.1.5 Slipids (Stockholm Lipids)

This force field that is applied to all-atomic has been extended to polyunsaturated lipids for simulations on lipid bilayers (Sonne et al. 2007). Jämbeck et al. depicted that the Slipids force field is a favorable set of parameters for the ab initio measurement of the force-field parameter (Jämbeck and Lyubartsev 2012). Atomic charges and torsion angles associated with polyunsaturated lipid tails have been parametrized using molecular structures. Bilayer simulations comprised of seven polyunsaturated lipids verified the new parameters of the area of power. It was noticed with the available data on the areas per lipids, bilayer volumetric properties, deuterium order parameters, and form factor scattering.

The potential function of all common force field divides into two categories. The first category, *bonded interactions*, includes in covalent bond-stretching, angle-bending, torsion potentials while rotating around bonds, and potential for "improper torsion" out-of-plane: the remaining category, *nonbonded interactions*, consists of Lennard-Jones repulsion and dispersion as well as Coulomb electrostatics (Dubbeldam et al. 2019).

9.3.2 Energy Minimization

The potential energy for the different interactions is estimated for one single conformation. Energy minimization is usually done through gradient optimization: atoms are moved to reduce the net forces correspond to the temperature (Adcock and McCammon 2006). The energy-reduced structure has minimal effects on each atom and therefore serves as an excellent point of departure for simulation. The energy minimization of the protein structure generated by MD simulation can be used to highlight the features of the energy landscape. The relative motion of structures provides an estimate of the size of features on the protein energy landscape. This method offers more knowledge on the lower energy surface potential than on the energy constraints between those minima. It is necessary to note that the free energies correlated with these substances, and their transitions are accessible only if sufficient barrier crossings are observed in the simulation (Troyer and Cohen 1995). The method is divided into subclasses by order of the derivative used to locate a minimum on the surface of the energy. The steepest descent and the conjugate gradient algorithm are available methods for minimized the small force on each atom

of the structure. The steepest descent method measured the energy for the initial geometry, and the atoms have been moved in a small increment in one of the coordinate system directions (Kraft 2017). The procedure has repeated for all atoms, which are eventually moved downhill on the energy surface to a new position. The conjugate gradient method assembles the function details from one iteration to the next. The opposite of the improvement made in an earlier iteration can be avoided with this technique. The gradient is determined for each minimization phase and used as additional information for the measurement of the minimization procedure's new path vector.

9.3.3 Conformational Search Algorithm

9.3.3.1 Global Optimization Algorithms

Optimization refers to find the minimum global energy of a potential surface. Due to the complicated existence of the potential energy functions of polypeptides, identifying potential energy minima is an extremely nonlinear and very complex problem of global optimization (Yurtkuran 2019). Several different methods have been developed, including predetermined branch-and-bound, probabilistic brute-force sampling, and pivot algorithms to solve this form of the global optimization problem.

9.3.3.2 Genetic Algorithm

Some of the most common approaches are geared toward using genetic algorithms (GA). GA and evolutionary computation are approaches that use computational evolution processes to solve complex problems. Unlike most other problem-solving methods, they work on a pool of individuals, each of which is a solution candidate with an associated fitness that gives a quality measure and enables one to rank and compare the solutions with each other (Hackenberger 2019).

9.3.3.3 Simulated Annealing

A technique with a predictive combination of simulated annealing-MD (SA-MD) and empirical-based peptide screening with high precision, unfolded, called as multiple simulated annealing-MD (MSA-MD) (Hollingsworth and Dror 2018). Through a large-scale sampling structure, we can detect a wider conformational space of a peptide or mini protein based on MSA-MD, and the near-native peptides and protein conformations can be acquired. This technique is applied to predict the structure of ALPHA1, Trp-cage protein, PolyAla, two peptides containing β sheet structure, and two mini-proteins containing over 40 residues (Hao et al. 2015).

9.3.3.4 Tabu Search

Tabu search (TS) is a metaheuristic technique designed to direct optimization methods in order to prevent local optimization during complex or multimodal numerical optimization problems (Glover 1986, 1990).TS method uses a variety of versatile memory cycles, each with different associated time scales, to allow more thorough use of search information than rigid memory or memory-less systems.

9.3.3.5 Lennard-Jones Potential

Hooke's law commonly meant to describe bonded interactions, and nonbonded atoms may be viewed as inelastic hard spheres or may interact with possible Lennard-Jones (Adcock and McCammon 2006). With these basic models, a simulation of MD numerically incorporates Newton's equations of motion, thus allowing for the measurement of structural variations in time.

9.4 Trajectory Analysis

9.4.1 Root Mean Square Deviation

The MD simulation trajectory involves simulation results, using which several properties including root mean square deviation (RMSD), root mean square fluctuation (RMSF), gyration radius, free-energy landscape, and principal component analysis (PCA) can be calculated which determine the consistency of the structure of a protein (Mehra et al. 2020; Shukla et al. 2018a, b, c, d, e). RMSD is the average displacement of the atoms relative to a reference structure at the moment in simulation (Fig. 9.3). This is useful for the study of the model structure's time-dependent motions (Kufareva and Abagyan 2012; Vijayakumar et al. 2018). RMSD is also used to determine if a system is stable in the timescale of the simulations or whether it diverges from the initial coordinates. The RMSD value represents the mean distance between the atoms (usually the backbone atoms) of the superimposed proteins. It can be extended to other molecules with no protein, such as small organic molecules. To calculate the RMSD, issue "*g_rms*" command in gromacs is as follows:

```
g_rms -f input.xtc -s input.pdb -o rmsd.xvg
```

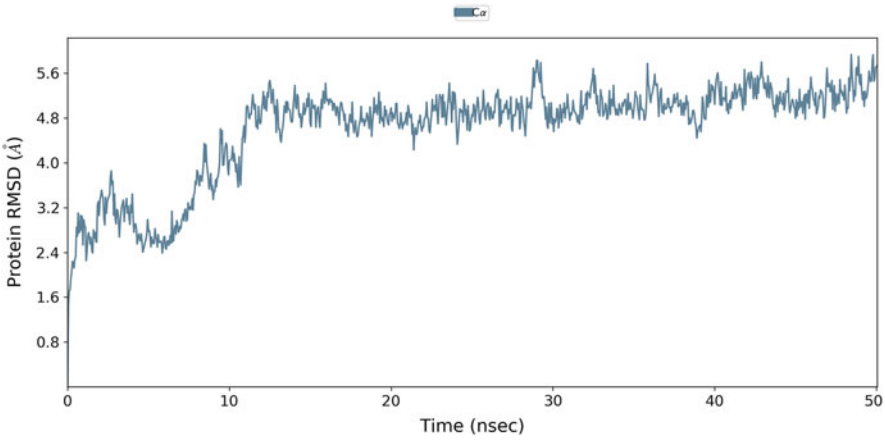


Fig. 9.3 RMSD plot of backbone atoms for the model system

g_rms measures the comparison between two structures by computing the RMSD with each structure in the trajectory (-f) compared with a reference in the structure file (-s).

9.4.2 Root Mean Square Fluctuation

Protein flexibility is essential for protein's function and ligand-binding procedure. MD simulation evolved into understanding the motions of macromolecular systems of high complexity and structure–function relationship (Hospital et al. 2015; Shukla et al. 2017a, b). The important part of the protein analysis consists of the description of the structural fluctuations of the macromolecule, which can be complex and hard to interpret from a functional perspective. RMSF is one of the most common measures of structural fluctuations. The RMSF is a proportion of the displacement of an individual atom or group of atoms, comparative with the reference structure, averaged over the number of atoms (Fraccalvieri et al. 2011) (Fig. 9.4). To calculate the RMSF, issue "*g_rmsf*" command in gromacs is as follows:

```
g_rmsf -f input.xtc -s input.pdb -o rmsf.xvg
```

"*g_rmsf*" command calculates the RMSF of atomic positions after adjustment to a reference frame.

9.4.3 Radius of Gyration

The radius of gyration (Rg) is an indicator of protein structure compactness. It is a concern with how secondary structures can compactly fold into protein's 3D

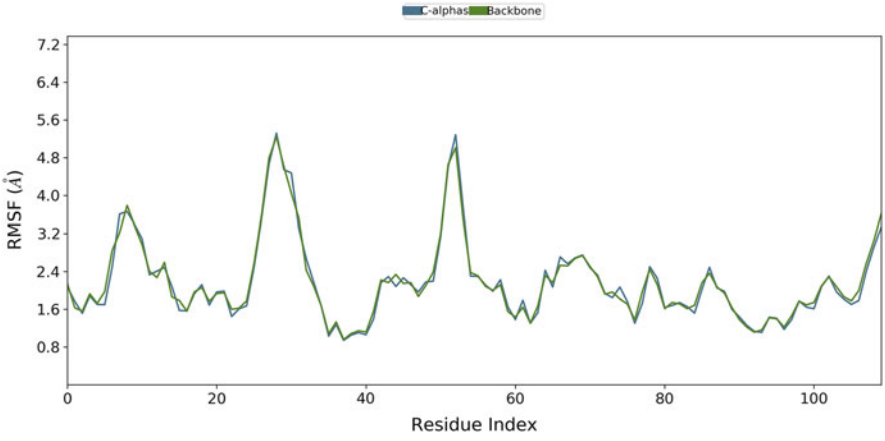


Fig. 9.4 The RMSF graph calculates the structural fluctuation position of protein atoms or residues. The green and blue lines represent the backbone and Cα RMSF, respectively

structure (Ivankov et al. 2009; Pandey et al. 2017; Shukla et al. 2017a, b). The correlation between folding and parameters that denote the protein complexity describes the compactness of the protein structure. "*g_gyrate*" command computes Rg of each atom of the protein model system. The atoms are explicitly mass-weighted.

```
g_gyrate -f input.xtc -s input.pdb -o gyrate.xvg
```

9.4.4 Free-Energy Landscape

The free-energy system (FEL) analyzes changes in the conformation of proteins. This approach is obtained by estimating the joint distribution of probabilities from the critical plane of the top two eigenvectors (Frauenfelder et al. 1991). The initial few eigenvectors represent the most dominant collective motions providing a useful definition of sub-conformational structural transitions (Sang et al. 2017).

The FEL underlies the thermodynamic functions of the protein complex that provide detailed information on the characterization of the amino acid interaction, which is relevant to the protein folding and binding. The FEL of the model pKID−KIX system (pKID, phosphorylated kinase-inducible domain; KIX, kinase interacting domain) was studied to represent energy landscape during simulation (Chong and Ham 2019). This system was measured by FEL that defines the landscape, which is based on fully atomistic, explicit-water with an aggregated simulation time of >30 μs. In this study, Chong et al. found that the landscape slope measuring the strength of the energetic bias against the protein folded state. Before binding to the other domain, pKID considerably unfolded, and after binding, the protein autonomously folds. The landscape figure indicates that the particular pKID remains disorder and not allow for folding before binding, and it became more progressively funneled as the domain bind with the KIX domain and adopts the folded structure (Chong and Ham 2019). The free energies of state A and B are defined as follows:

$$F_{\mathrm{A,B}} = -1/\beta \ \ln \ \mathrm{ZA,B},$$

A region represents a fluctuating structure of a protein on a high-dimensional complex free-energy landscape, which dynamically explored (Chipot and Pohorille 2007). Go-kit (Neelamraju et al. 2019), GROMACS suite (Lindahl et al. 2001), and PLUMED portable plugin (Bonomi et al. 2009) are such packages for the FEL calculations. An energy landscape for binding of a ligand with the target site of protein is shown in Fig 9.5a, which represents different levels of energy for a protein before binding, during binding, and after binding of a ligand. The energy landscape for different conformation of a protein during the folded and unfolded state is shown in Fig. 9.5b, which indicates the low free energy for folded states while very high free energy for unfolded states of a protein.

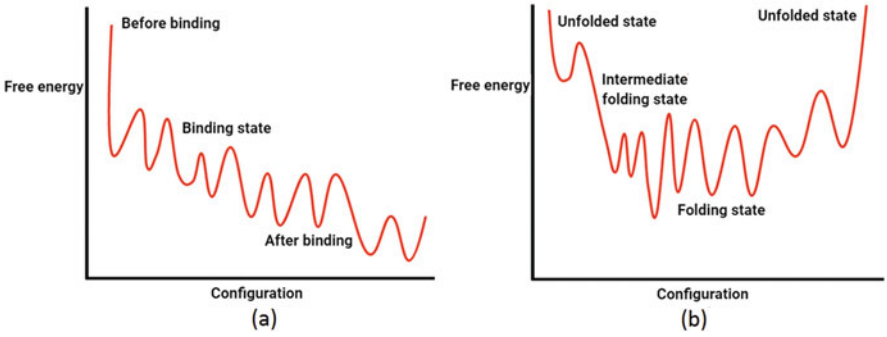


Fig. 9.5 Free-energy landscape of a protein. (**a**) Binding of ligand with a protein. (**b**) Protein in unfolded and folded state

9.4.5 Free-Energy Binding

Binding efficiency characterizes the strength of protein–protein, protein–ligand, and protein–peptide docking (Brandsdal et al. 2003; Gilson and Zhou 2007; Shukla et al. 2018a, b, c, d, e). The advancement in computing, prediction of binding efficiency of such recognition is based on MM principles (Du et al. 2016). The relative or absolute binding-free energies analyzed by several methods cover a broad range of accuracies and computational requirements. The molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) model also known as an end-point free-energy model as it determines the free energy changes by the evaluation of the first and last state of protein-ligand systems (Homeyer and Gohlke 2012). An end-point method is computationally powerful and thus are commonly discussed and applied in the literature. The accuracy of the MM/PBSA depends on the connection that has been established between statistical thermodynamics and various endpoint free-energy models. There are certain limitations of this method, which include the fact that there is no regard for particular water interactions, insensitive to the trajectory, and sensitive to induced fit effects (Wong et al. 2009).

MM/PBSA is essentially a post-analysis approach to assess the free molecular energies or the binding free energies of molecular complexes (Genheden and Ryde 2015; Shukla et al. 2018a, b, c, d, e). The approach divides the free energy into molecular mechanical energies, continuum solvation energies, and solute entropy terms as given below.

$$\Delta \text{Gbinding} = \text{Gcomplex} - (\text{Grecptor} + \text{Gligand})$$

where Gcomplex indicates the total free energy upon protein–protein complex and Grecptor and Gligand are total free energies of the receptor and ligand in a solvent, respectively.

For the estimation of the free energies, various robust methods have been developed which include the linear interaction energy (LIE) method (Gutiérrez-de-Terán and Aqvist 2012), the chemical Monte Carlo/MD (CMC/MD) method

(Eriksson et al. 1999), the pictorial representation of free-energy components (PRO-FEC) method (Wang et al. 1998), the one-window free-energy grid (OWFEG) method (Pearlman 1999), the λ-dynamics method, and the 4D-PMF method.

9.4.6 Principal Component Analysis

MD simulation approach generates trajectories to understand the folding/unfolding characteristics and conformation of proteins best represented by a large number of dimensions (David and Jacobs 2014; Kalita et al. 2017; Shukla et al. 2018a, b, c, d, e). The detailed study of trajectories at different temperatures need implementation of the statistical method, such as principal component analysis (PCA). A successful multivariate technique used to characterize the protein conformation by reducing or simplifying large and complicated datasets (Jolliffe and Cadima 2016; David and Jacobs 2014). In 1901, Karl Pearson first introduced the PCA approach for just a few variables (Pearson 1901). Many variants have been proposed and introduced, such as the critical method of dynamics, which was commonly used and published in the recent literature. However, the underlying mathematical procedure for essential dynamics (ED) remains the same as PCA. It is a covariance-matrix-based technique that reduces the multi-dimensional protein complex to a lower dimension, along which the diffusive properties can be identified at all the stages of protein folding (Maisuradze et al. 2009).

PCA is a multivariate statistical technique, which aims to reduce the high-dimensional space of variables to a lower one (Jolliffe and Cadima 2016). When it comes to MD simulations of proteins, in its simplest sense, it may assume that each snapshot from the simulation is a sample conformation from the equilibrated fluctuations of a protein (Salsbury 2010). Furthermore, it assumes that samples obtained from the simulation analysis span most of the ED of the protein (David and Jacobs 2014; David et al. 2017). In short, the use of PCA analysis to identify the most significant protein motions has become commonplace. MDWeb (Hospital et al. 2012), pyPcazip (Shkurti et al. 2016), JED (David et al. 2017), Wordom (Seeber et al. 2007), and NAMD (Phillips et al. 2005) packages eventually provide PCA to examine protein trajectories; researchers frequently make inferences of their observations without insight into how to make interpretations, and they are also unaware of the shortcomings and generalizations of such study.

9.5 Softwares for MD Simulation

MD software packages are available for studying the protein, nucleic acids, and carbohydrate dynamics. Different packages have different features and merits, few are freely available, and few are commercial. A brief description is tabulated in Table 9.2.

Table 9.2 The application-based MD simulation software is listed as a tabulated form

Software name	Application	License	References
Abalone (http://www.biomolecular-modeling.com/Abalon)	Biomolecular simulations, protein folding.	Proprietary, gratis, commercial	Elmore (2016); Spitznagel et al. (2016)
ABINIT (http://www.abinit.org)	Calculate the total energy, charge density, and electronic structure of molecules and periodic solids with density functional theory (DFT) and many-body perturbation theory (MBPT), using pseudopotentials and a plane-wave or wavelet basis. ABINIT also can optimize the geometry, perform MD simulations, or generate dynamical matrices, born effective charges, and dielectric tensors, and many more properties.	GPL	Pirhadi et al. (2016)
ADF (https://www.scm.com/doc/QMMM/ADF_QMMM/ADF_QMMM.htmL)	Modeling suite: ReaxFF, UFF, QM-MM with Amber and Tripos force fields, DFT and semiempirical methods, conformational analysis with RDKit; partly GPU-accelerated	Proprietary, commercial, gratis trial	Klamt (2005); Young (2001)
Ascalaph designer (http://www.biomolecular-modeling.com/Ascalaph/Ascalaph_Designer.html)	Molecular building (DNA, proteins, hydrocarbons, nanotubes), MD, GPU acceleration	Mixed: free open-source (GNU GPL) & commercial	Loukatou et al. (2014)
CHARMM (https://www.charmm.org/)	A commercial version with multiple graphical front ends is sold by Accelrys (as CHARMm)	Proprietary, commercial	Karplus (2006)
CP2K (https://www.cp2k.org)	CP2K can perform atomistic and molecular simulations of solid-state, liquid, and biological systems.	Free, open-source GNU GPLv2 or later	Laino et al. (2005); Kühne (2014)
Dacapo (https://wiki.fysik.dtu.dk/dacapo)	Total energy program that uses density functional theory. It can do MD/structural	–	Bahn and Jacobsen (2002)

(continued)

Table 9.2 (continued)

Software name	Application	License	References
	relaxation while solving the Schrödinger equations. It has support for parallel execution and is used through the atomic simulation environment (ASE)		
Desmond (https://www.schrodinger.com/desmond)	High-performance MD has comprehensive GUI to build, visualize, and review results and calculation setup and launch	Proprietary, commercial or gratis	Jorgensen (2009); Desmond (2012); Ivanova et al. (2018)
Discovery studio (https://www.3dsbiovia.com)	Comprehensive life science modeling and simulation suite of applications focused on optimizing drug discovery process: small molecule simulations, QM-MM, pharmacophore modeling, QSAR, protein–ligand docking, protein homology modeling, sequence analysis, protein–protein docking, antibody modeling, etc.	Proprietary, trial available	Accelrys Software Inc. (2012)
GROMACS (www.gromacs.org)	High-performance MD	Free open-source GNU GPL	Lindahl et al. (2001)
GROMOS(http://www.gromos.net)/	Intended for biomolecules	Proprietary, commercial	Pol-Fachin et al. (2012)
HyperChem (http://www.hyper.com)	MM+, Ambers, Amber2, Amber3, Amber94, Amber96, Amber99, Bio + 83, Bio + 85, Charmm-19, Charmm-22, Charmm-27, OPLS, Custom, Extended-Hukel, CNDO, INDO, MINDO3, MNDO, MNDO/d, AM1, PM3, RM1, ZINDO/1, ZINDO/s, TNDO, Hartreee-Fock, MP2, CI, density functional theories, solvent model, conformational	Proprietary, trial available	Hypercube (2002)

(continued)

Table 9.2 (continued)

Software name	Application	License	References
	sampling, minimizing, MD, MC, Langevin. QM/MM calculations can be used. Includes the HyperChem GUI, which provides visualizing, molecule building (polypeptide (protein), nucleic acid, polysaccharides, polymer builders), calculation setup, job launch, and monitoring, project-level organizing of results, access to a suite of other modeling programs. Developer's kit using HCL, C, C++, Visual Basic, Fortran, Tcl/Tk, DDE.		
LAMMPS (https://lammps.sandia.gov)	Has potentials for soft and solid-state materials and coarse-grain systems	Free open-source, GNU GPLv2	Grindon et al. (2004)
MacroModel (https://www.schrodinger.com/macromodel)	OPLS-AA, MMFF, GBSA solvent model, conformational sampling, minimizing, MD. Includes the Maestro GUI, which provides visualizing, molecule building, calculation setup, job launch and monitoring, project-level organizing of results, access to a suite of other modeling programs.	Proprietary	Watts et al. (2014)
MAPS (https://www.scienomics.com)	Building, visualizing, and analysis tools in one user interface, with access to multiple simulation engines	Proprietary, trial available	Yao et al. (2018)
Materials studio (https://www.3dsbiovia.com/portfolio/materials-studio.html)	An environment that brings materials simulation technology to desktop computing, solving key problems in R&D processes	Proprietary, trial available	Sharma (2019)

(continued)

Table 9.2 (continued)

Software name	Application	License	References
MBN ExplorerMBN studio (https://quantbiolab.com/research/mbn-explorer)	Standard and reactive CHARMM force fields; molecular modeler (carbon nanomaterials, biomolecules, nanocrystals); an explicit library of examples	Proprietary, free trial available	Sushko et al. (2019)
MDynaMix (http://www.fos.su.se/~sasha/mdynamix)	Parallel MD	Free, open-source GNU GPL	Lyubartsev and Laaksonen (2000)
MOE (https://www.chemcomp.com)	Molecular operating environment (MOE)	Proprietary	Molecular Operating Environment (MOE) (2016)
ORAC (https://brehm-research.de/orcamd.php)	MD simulation program to explore free-energy surfaces in biomolecular systems at the atomic level	Free open source	Marsili et al. (2010)
NAMD + VMD (https://www.ks.uiuc.edu/Research/namd)	Fast, parallel MD, CUDA	Proprietary, free academic use, source code	Stone et al. (2016)
NWChem (http://www.nwchem-sw.or)	High-performance computational chemistry software, includes quantum mechanics, MD, and combined QM-MM methods	Free open-source, educational community license version 2.0	Straatsma and McCammon (2001)
Protein local optimization program (http://www.jacobsonlab.org/plop_manual/plop_overview.htm)	Helix, loop, and side-chain optimizing, fast energy minimizing	Proprietary	Jacobson et al. (2002)
Q (http://xray.bmc.uu.se/~aqwww/q)	(1) Free-energy perturbation (FEP) simulations, (2) empirical valence bond (EVB), calculations of reaction-free energies, (3) linear interaction energy (LIE) calculations of receptor–ligand binding affinities	Uppsala Molekylmekaniska HB	Marelius et al. (1998)
QuantumATK (https://www.synopsys.com/silicon/quantumatk.html)	A complete atomistic modeling platform for material science. It includes DFT (plane-wave and LCAO), semiempirical, and force-field simulation engines.	Proprietary, commercial	Smidstrup et al. (2019)

(continued)

Table 9.2 (continued)

Software name	Application	License	References
QUANTUM ESPRESSO (https://www.schrodinger.com/products/quantum-espresso)	Designed for modeling at the nanoscale using DFT, plane waves, and pseudopotentials and its capabilities include ground-state calculations, structural optimization, transition states, and minimum energy paths, ab initio MD, DFT perturbation theory, spectroscopic properties, and quantum transport.		Ohto et al. (2019)
RMG (http://rmgdft.sourceforge.net)	DFT code that uses real-space grids to provide high scalability across thousands of processors and GPU acceleration for both structural relaxation and MD.		Moore et al. (2002)
SAMSON (https://www.samson-connect.net)	Computational nanoscience (life sciences, materials, etc.). Modular architecture, modules termed SAMSON elements	Proprietary, gratis	Jaillet et al. (2017)
Scigress (http://www.scigress.com)	MM, DFT, semiempirical methods, parallel MD, conformational analysis, linear scaling SCF, docking protein–ligand, batch processing, virtual screening, automated builders (MD, proteins, crystals)	Proprietary	Marchand et al. (2014)
Siam quantum (https://sites.google.com/site/siamquantum)	Optimized for parallel computation and its capabilities include the calculation of Hartree–Fock and MP2 energies, minimum energy crossing point calculations, geometry optimization, population analysis, and quantum MD.	–	Djidjev et al. (2019)

(continued)

Table 9.2 (continued)

Software name	Application	License	References
TeraChem (http://petachem.com)	High-performance GPU-accelerated ab initio MD and TD/DFT software package for very large molecular or even *nanoscale* systems. Runs on NVIDIA GPUs and 64-bit Linux has heavily optimized CUDA code.	Proprietary, trial licenses available	Luehr et al. (2015)
TINKER (http://tinkertools.org)	Software tools for molecular design-Tinker-OpenMM	Proprietary, gratis	Rackers et al. (2018)
	Software tools for molecular design-Tinker-HP		
Tremolo-X (http://www.tremolo-x.com)	Fast, parallel MD	Proprietary	Dolado et al. (2010)
YASARA (http://www.yasara.org)	Molecular modeling, docking, graphics, and MD simulation	Proprietary	Land and Humble (2018)

9.6 Conclusion

MD simulations have become useful methods for sampling protein structure information and dynamics. It gives insights into the internal motions of proteins and the structure of protein complexes. Recently, further emphasis on the system and techniques has underpinned the advancement of force fields which are suitable for the simulations with a wide range of drug molecules. The general approach to characterize the protein conformations and determine the free energies from MD simulations, without the requirement for reasonably clustering methods, while not disregarding the multidimensional configuration space in which protein movements are still a challenge. It is important to tackle this challenge to develop conformationally selective drugs. Besides, the investigation of protein dynamics is often semi-quantitative while associated motion matrices or any desired correlation function can easily be calculated quantitatively and qualitatively. The main challenges are the amount of knowledge that can be obtained from the MD simulations experiment. It is possible to make hopefully continuous progress in applying statistical methods and analyses that incorporate structural and dynamic knowledge to enhance the quantitative and routine analysis of simulation data.

References

Aamir M, Singh VK, Dubey MK, Meena M, Kashyap SP, Katari SK, Upadhyay RS, Umamaheswari A, Singh S (2018) In silico prediction, characterization, molecular docking, and dynamic studies on fungal SDRS as novel targets for searching potential fungicides against fusarium wilt in tomato. Front Pharmacol 9:1038

Accelrys Software Inc (2012) Discovery studio, release 3.0. Accelrys Inc., San Diego, CA. www.accelrys.com

Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev 106:1589–1615

Alder BJ, Wainwright TE (1959) Studies in molecular dynamics I. General method. J Chem Phys 31:459–466

Alder BJ, Wainwright TE (1960) Studies in molecular dynamics. II. Behavior of a small number of elastic spheres. J Chem Phys 33:1439–1451

Bahn S, Jacobsen K (2002) An object-oriented scripting interface to a legacy electronic structure code. Comput Sci Eng 4:56–66

Beck DA, Daggett V (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. Methods 34:112–120

Bonomi M, Branduardi D, Bussi G, Camilloni C, Provasi D, Raiteri P, Donadio D, Marinelli F, Pietrucci F, Broglia RA, Parrinello M (2009) PLUMED: a portable plugin for free-energy calculations with molecular dynamics. Comput Phys Commun 180:1961–1972

Brandsdal BO, Osterberg F, Almlöf M, Feierberg I, Luzhkov VB, Aqvist J (2003) Free energy calculations and ligand binding. Adv Protein Chem 66:123–158

Breda A, Valadares NF, de Souza ON, Garratt RC (2007) Protein structure, modelling and applications. In: Bioinformatics in tropical disease research: a practical and case-study approach. National Center for Biotechnology Information, Bethesda

Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30:1545–1614

Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688

Chipot C, Pohorille A (2007) Free energy calculations. In: Springer series in chemical physics. Springer, Berlin

Chong SH, Ham S (2019) Folding free energy landscape of ordered and intrinsically disordered proteins. Sci Rep 9:1–9

Collier TA, Piggot TJ, Allison JR (2020) Molecular dynamics simulation of proteins. Methods Mol Biol 2073:311–327

David CC, Jacobs DJ (2014) Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol 1084:193–226

David CC, Singam ERA, Jacobs DJ (2017) JED: a Java essential dynamics program for comparative analysis of protein trajectories. BMC Bioinf 18:271

Desmond Molecular Dynamics System, version 3.1 (2012) DE Shaw Research, New York, Maestro–Desmond Interoperability Tools, version 3.1, 2012, Schrödinger, New York

Djidjev HN, Hahn G, Mniszewski SM, Negre CFA, Niklasson AMN (2019) Using graph partitioning for scalable distributed quantum molecular dynamics. Algorithms 12:187

Dolado JS, Griebel M, Hamaekers J, Heber F (2010) The nano-branched structure of cementitious calcium-silicate-hydrate gel. J Mater Chem A 21:4445–4449

Dror RO, Jensen MØ, Borhani DW, Shaw DE (2010) Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. J Gen Physiol 135:555–562

Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, Ji XL, Liu SQ (2016) Insights into protein-ligand interactions: mechanisms, models, and methods. Int J Mol Sci 17:144

Dubbeldam D, Walton KS, Vlugt TJ, Calero S (2019) Design, parameterization, and implementation of atomic force fields for adsorption in nanoporous materials. Adv Theory Simul 2:1900135. https://doi.org/10.1002/adts.201900135
Elmore DE (2016) Why should biochemistry students be introduced to molecular dynamics simulations—and how can we introduce them? Biochem Mol Biol Educ 44:118–123
Eriksson MA, Pitera J, Kollman PA (1999) Prediction of the binding free energies of new TIBO-like HIV-1 reverse transcriptase inhibitors using a combination of PROFEC, PB/SA, CMC/MD, and free energy calculations. J Med Chem 42:868–881
Forster MJ (2002) Molecular modelling in structural biology. Micron 33:365–384
Fraccalvieri D, Pandini A, Stella F, Bonati L (2011) Conformational and functional analysis of molecular dynamics trajectories by self-organising maps. BMC Bioinf 12:158
Frauenfelder H, Sligar S, Wolynes P (1991) The energy landscapes and motions of proteins. Science 254:1598–1603
Geng H, Chen F, Ye J, Jiang F (2019) Applications of molecular dynamics simulation in structure prediction of peptides and proteins. Comput Struct Biotechnol J 17:1162–1170
Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discovery 10:449–461
Gentle JE (2009) Computational statistics. Springer, New York
Gilson MK, Zhou HX (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
Glover F (1986) Future paths for integer programming and links to artificial intelligence. Comput Oper Res 13:533–549
Glover F (1990) Tabu search—part 2. ORSA J Comput 2:4–32
Grindon C, Harris S, Evans T, Novik K, Coveney P, Laughton C (2004) Large-scale molecular dynamics simulation of DNA: implementation and validation of the AMBER98 force-field in LAMMPS. Philos Transact A Math Phys Eng Sci 362:1373–1386
Gutiérrez-de-Terán H, Aqvist J (2012) Linear interaction energy: method and applications in drug design. Methods Mol Biol 819:305–323
Guvench O, MacKerell AD Jr (2008) Comparison of protein force fields for molecular dynamics simulations. Methods Mol Biol 443:63–88
Hackenberger BK (2019) Genetics without genes: application of genetic algorithms in medicine. Croat Med J 60:177
Hao GF, Xu WF, Yang SG, Yang GF (2015) Multiple simulated annealing-molecular dynamics (msa-md) for conformational space search of peptide and miniprotein. Sci Rep 5:15568. https://doi.org/10.1038/srep15568
Hernández-Rodríguez M, Rosales-Hernández MC, Mendieta-Wejebe JE, Martínez-Archundia M, Basurto JC (2016) Current tools and methods in molecular dynamics (MD) simulations for drug design. Curr Med Chem 23:3909–3924
Hollingsworth SA, Dror RO (2018) Molecular dynamics simulation for all. Neuron 99:1129–1143
Homeyer N, Gohlke H (2012) Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. Mol Inf 31:114–122
Hospital A, Andrio P, Fenollosa C, Cicin-Sain D, Orozco M, Gelpí JL (2012) MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. Bioinformatics 28:1278–1279
Hospital A, Goñi JR, Orozco M, Gelpí JL (2015) Molecular dynamics simulations: advances and applications. Adv Appl Bioinforma Chem 8:37
Hubbard D, Samuelson DA (2009) Modeling without measurements: how the decision analysis culture's lack of empiricism reduces its effectiveness. OR/MS Today 36:26–31
Hug S (2013) Classical molecular dynamics in a nutshell. Methods Mol Biol 924:127–152
Hypercube (2002) HyperChem 7.52: molecular visualization and simulation program package. Hyperchbe, Gainsville, FL
Ingólfsson HI, Arnarez C, Periole X, Marrink SJ (2016) Computational 'microscopy' of cellular membranes. J Cell Sci 129:257–268

Ivankov DN, Bogatyreva NS, Lobanov MY, Galzitskaya OV (2009) Coupling between properties of the protein shape and the rate of protein folding. PLoS One 4(8):e6476

Ivanova L, Tammiku-Taul J, García-Sosa AT, Sidorova Y, Saarma M, Karelson M (2018) Molecular dynamics simulations of the interactions between glial cell line-derived neurotrophic factor family receptor GFRα1 and small-molecule ligands. ACS Omega 3:11407–11414

Jacobson MP, Kaminski GA, Friesner RA, Rapp CS (2002) Force-field validation using protein side chain prediction. J Phys Chem B 106:11673–11680

Jaillet L, Artemova S, Redon S (2017) IM-UFF: extending the universal force-field for interactive molecular modeling. J Mol Graph Model 77:350–362

Jämbeck JP, Lyubartsev AP (2012) Derivation and systematic validation of a refined all-atom force-field for phosphatidylcholine lipids. J Phys Chem B 116:3164–3179

Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Philos Trans R Soc A Math Phys Eng Sci 374:20150202

Jorgensen WL (2009) Efficient drug lead discovery and optimization. Acc Chem Res 42:724–733

Kalita J, Shukla R, Shukla H, Gadhave K, Giri R, Tripathi T (2017) Comprehensive analysis of the catalytic and structural properties of a mu-class glutathione s-transferase from *Fasciola gigantica*. Sci Rep 7:17547

Kalita P, Shukla H, Gadhave K, Giri R, Tripathi T (2018) Role of the glutaredoxin domain and FAD in the stabilization of thioredoxin glutathione reductase. Arch Biochem Biophys 656:38–45

Kalita J, Shukla R, Tripathi T (2019a) Structural basis of urea-induced unfolding of *Fasciola gigantica* glutathione S-transferase. J Cell Physiol 234(4):4491–4503

Kalita P, Das KC, Shukla H, Tripathi T (2019b) Conserved Arg451 residue is critical for maintaining the stability and activity of thioredoxin glutathione reductase. Arch Biochem Biophys 674:108098

Kantarci-Carsibasi N, Haliloglu T, Doruker P (2008) Conformational transition pathways explored by Monte Carlo simulation integrated with collective modes. Biophys J 95:5862–5873

Karplus M (2006) Spinach on the ceiling: a theoretical chemist's return to biology. Annu Rev Biophys Biomol Struct 35:1–47

Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9:646–652

Klamt A (2005) COSMO-RS from quantum chemistry to fluid phase thermodynamics and drug design, 1st edn. Elsevier, Amsterdam

Kraft D (2017) Self-consistent gradient flow for shape optimization. Optim Methods Softw 32:790–812

Kufareva I, Abagyan R (2012) Methods of protein structure comparison. Methods Mol Biol 857:231–257

Kühne TD (2014) Second generation Car–Parrinello molecular dynamics. WIRES Comput Mol Sci 4:391–406

Laino T, Mohamed F, Laio A, Parrinello M (2005) An efficient real space multigrid QM/MM electrostatic coupling. J Chem Theory Comput 1:1176–1184

Land H, Humble MS (2018) YASARA: a tool to obtain structural guidance in biocatalytic investigations. Methods Mol Biol 1685:43–67

Lee EH, Hsin J, Sotomayor M, Comellas G, Schulten K (2009) Discovery through the computational microscope. Structure 17:1295–1306

Levitt M, Lifson S (1969) Refinement of protein conformations using a macromolecular energy minimization procedure. J Mol Biol 46:269–279

Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7:306–317

Lorenz C, Doltsinis NL (2012) Molecular dynamics simulation: from "ab initio" to "coarse grained". In: Leszczynski J (ed) Handbook of computational chemistry. Springer, Dordrecht, pp 195–238

Loukatou S, Papageorgiou L, Fakourelis P, Filntisi A, Polychronidou E, Bassis I, Megalooikonomou V, Makałowski W, Vlachakis D, Kossida S (2014) Molecular dynamics simulations through GPU video games technologies. J Mol Biochem 3:64

Luehr N, Jin AG, Martínez TJ (2015) Ab initio interactive molecular dynamics on graphical processing units (GPUs). J Chem Theory Comput 11:4536–4544

Lyubartsev AP, Laaksonen AM (2000) Dyna Mix–a scalable portable parallel md simulation package for arbitrary molecular mixtures. Comput Phys Commun 128:565–589

Macchiagodena M, Del Frate G, Brancato G, Chandramouli B, Mancini G, Barone V (2017) Computational study of the DPAP molecular rotor in various environments: from force-field development to molecular dynamics simulations and spectroscopic calculations. Phys Chem Chem Phys 19:30590–30602

Maisuradze GG, Liwo A, Scheraga HA (2009) Principal component analysis for protein folding dynamics. J Mol Biol 385:312–329

Marchand N, Lienard P, Siehl HU, Izato H (2014) Applications of molecular simulation software SCIGRESS in industry and university. FUJITSU Sci Tech J 50:46–51

Marelius J, Kolmodin K, Feierberg I, Aqvist J (1998) Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. J Mol Graph Model 16:213–225

Marsili S, Signorini GF, Chelli R, Marchi M, Procacci P (2010) ORAC: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. J Comput Chem 31:1106–1116

McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590

Mehra R, Dehury B, Kepp KP (2020) Cryo-temperature effects on membrane protein structure and dynamics. Phys Chem Chem Phys 22:5427–5438

Molecular Operating Environment (MOE) (2016) Chemical Computing Group Inc., Montreal, QC, Canada

Moore S, Briggs E, Hodak M, Lu W, Bernholc J, Lee CW (2002) Scaling the RMG quantum mechanics code. In: Proceedings of the Extreme Scaling Workshop, vol 8, pp 1–6

Neelamraju S, Wales DJ, Gosavi S (2019) Go-Kit: a tool to enable energy landscape exploration of proteins. J Chem Inf Model 59:1703–1708

Ohto T, Dodia M, Xu J, Imoto S, Tang F, Zysk F, Kühne TD, Shigeta Y, Bonn M, Wu X, Nagata Y (2019) Accessing the accuracy of density functional theory through structure and dynamics of the water–air interface. J Phys Chem Lett 10:4914–4919

Pandey T, Shukla R, Shukla H, Sonkar A, Tripathi T, Singh AK (2017) A combined biochemical and computational studies of the rho-class glutathione s-transferase sll1545 of *Synechocystis* PCC 6803. Int J Biol Macromol 94:378–385

Paquet E, Viktor HL (2015) Molecular dynamics, Monte Carlo simulations, and langevin dynamics: a computational review. Biomed Res Int 2015:183918

Pearlman DA (1999) Free energy grids: a practical qualitative application of free energy perturbation to ligand design using the OWFEG method. J Med Chem 42:4313–4324

Pearson K (1901) On lines and planes of closest fit to systems of points in space. Phil Mag 2:559–572

Pensak DA (1989) Molecular modelling: scientific and technological boundaries. Pure Appl Chem 61:601–603

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802

Pirhadi S, Sunseri J, Koes DR (2016) Open source molecular modeling. J Mol Graph Model 69:127–143

Pol-Fachin L, Rusu VH, Verli H, Lins RD (2012) GROMOS 53A6GLYC, an improved GROMOS force-field for hexopyranose-based carbohydrates. J Chem Theory Comput 8:4681–4690

Rackers JA, Wang Z, Lu C, Laury ML, Lagardère L, Schnieders MJ, Piquemal JP, Ren P, Ponder JW (2018) Tinker 8: software tools for molecular design. J Chem Theory Comput 14:5273–5289

Rahman A (1964) Correlations in the motion of atoms in liquid argon. Phys Rev 136:A405
Ryazantsev MN, Nikolaev DM, Struts AV, Brown MF (2019) Quantum mechanical and molecular mechanics modeling of membrane-embedded rhodopsins. J Membr Biol 252:425–449
Salmaso V, Moro S (2018) Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview. Front Pharmacol 9:923
Salsbury FR Jr (2010) Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. Curr Opin Pharmacol 10:738–744
Sang P, Du X, Yang LQ, Meng ZH, Liu SQ (2017) Molecular motions and free-energy landscape of serine proteinase K in relation to its cold-adaptation: a comparative molecular dynamics simulation study and the underlying mechanisms. RSC Adv 7:28580–28590
Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A (2007) Wordom: a program for efficient analysis of molecular dynamics simulations. Bioinformatics 23:2625–2627
Sharma S (2019) Molecular dynamics simulation of nanocomposites using BIOVIA materials studio, lammps and gromacs. Elsevier, Amsterdam
Shkurti A, Goni R, Andrio P, Breitmoser E, Bethune I, Orozco M, Laughton CA (2016) pyPcazip: a PCA-based toolkit for compression and analysis of molecular simulation data. SoftwareX 5:44–50
Shukla H, Shukla R, Sonkar A, Tripathi T (2017a) Alterations in conformational topology and interaction dynamics caused by L418A mutation leads to activity loss of *Mycobacterium tuberculosis* isocitrate lyase. Biochem Biophys Res Commun 490(2):276–282
Shukla H, Shukla R, Sonkar A, Pandey T, Tripathi T (2017b) Distant Phe345 mutation compromises the stability and activity of *Mycobacterium tuberculosis* isocitrate lyase by modulating its structural flexibility. Sci Rep 7:1058
Shukla R, Chetri PB, Sonkar A, Pakharukova MY, Mordvinov VA, Tripathi T (2018a) Identification of novel natural inhibitors of *Opisthorchis felineus* cytochrome P450 using structure-based screening and molecular dynamic simulation. J Biomol Struct Dyn 36(13):3541–3556
Shukla R, Shukla H, Tripathi T (2018b) Activity loss by H46 mutation in *Mycobacterium tuberculosis* isocitrate lyase is due to decrease in structural plasticity and collective motions of the active site. Tuberculosis 108:143–150
Shukla R, Shukla H, Kalita P, Tripathi T (2018c) Structural insights into natural compounds as inhibitors of *Fasciola gigantica* thioredoxin glutathione reductase. J Cell Biochem 119:3067–3080
Shukla R, Shukla H, Kalita P, Sonkar A, Pandey T, Singh DB, Kumar A, Tripathi T (2018d) Identification of potential inhibitors of *Fasciola gigantica* thioredoxin1: computational screening, molecular dynamics simulation and binding free energy studies. J Biomol Struct Dyn 36 (8):2147–2162
Shukla R, Shukla H, Sonkar A, Pandey T, Tripathi T (2018e) Structure-based screening and molecular dynamics simulations offer novel natural compounds as potential inhibitors of *Mycobacterium tuberculosis* isocitrate lyase. J Biomol Struct Dyn 36(8):2045–2057
Shukla R, Shukla H, Tripathi T (2019) Structural and energetic understanding of novel natural inhibitors of *Mycobacterium tuberculosis* malate synthase. J Cell Biochem 120:2469–2482
Singh S, Singh VK, Rai G (2020) Identification of differentially expressed hematopoiesis-associated genes in term low birth weight newborns by systems genomics approach. Current Genomics 20:469–482
Smith LG, Zhao J, Mathews DH, Turner DH (2017) Physics-based all-atom modeling of RNA energetics and structure. Wiley Interdiscip Rev RNA 8(5):10.1002/wrna.1422
Smidstrup S, Markussen T, Vancraeyveld P, Wellendorff J, Schneider J, Gunst T, Verstichel B, Stradi D, Khomyakov PA, Vej-Hansen UG, Lee ME (2019) QuantumATK: an integrated platform of electronic and atomic-scale modelling tools. J Phys Condens Matter 32:015901
Sonkar A, Shukla H, Shukla R, Kalita J, Tripathi T (2019) Unfolding of *Acinetobacter baumannii* MurA proceeds through a metastable intermediate: a combined spectroscopic and computational investigation. Int J Biol Macromol 126:941–951

Sonne J, Jensen MOØ, Hansen FY, Hemmingsen L, Peters GH (2007) Reparameterization of all-atom dipalmitoylphosphatidylcholine lipid parameters enables simulation of fluid bilayers at zero tension. Biophys J 92:4157–4167

Spitznagel B, Pritchett PR, Messina TC, Goadrich M, Rodriguez J (2016) An undergraduate laboratory activity on molecular dynamics simulations. Biochem Mol Biol Edu 44:130–139

Stone JE, Hynninen AP, Phillips JC, Schulten K (2016) Early experiences porting the NAMD and VMD molecular simulation and analysis software to GPU-accelerated OpenPOWER platforms. High Perform Comput 9945:188–206

Straatsma TP, McCammon JA (2001) IBM Syst J 40:328

Sushko GB, Solov'yov IA, Solov'yov AV (2019) ModelingMesoBioNano systems with MBN studio made easy. J Mol Graph Model 88:247–260

Sweere AJ, Fraaije JG (2017) Accuracy test of the OPLS-AA force-field for calculating free energies of mixing and comparison with PAC-MAC. J Chem Theory Comput 13:1911–1923

Troyer JM, Cohen FE (1995) Protein conformational landscapes: energy minimization and clustering of a long molecular dynamics trajectory. Proteins Struct Funct Genet 23:97–110

Vanommeslaeghe K, Guvench O (2014) Molecular mechanics. Curr Pharm Des 20:3281–3292

Vanommeslaeghe K, MacKerell AD Jr (2015) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochim Biophys Acta 1850:861–871

Vijayakumar R, Shukla R, Shukla H, Tripathi T (2018) Structure-function studies of the asparaginyl-tRNA synthetase from *Fasciola gigantica:* understanding the role of catalytic and non-catalytic domains. Biochem J 475(21):3377–3391

Wang L, Veenstra DL, Radmer RJ, Kollman PA (1998) Can one predict protein stability? An attempt to do so for residue 133 of T4 lysozyme using a combination of free energy derivatives, PROFEC, and free energy perturbation methods. Proteins 32:438–458

Watts KS, Dalal P, Tebben AJ, Cheney DL, Shelley JC (2014) Macrocycle conformational sampling with MacroModel. J Chem Inf Model 54:2680–2696

Wong S, Amaro RE, McCammon JA (2009) MM-PBSA captures key role of intercalating water molecules at a protein-protein interface. J Chem Theory Comput 5:422–429

Yao H, Dai Q, You Z, Bick A, Wang M (2018) Modulus simulation of asphalt binder models using molecular dynamics (MD) method. Constr Build Mater 162:430–441

Young DC (2001) Computational chemistry: a practical guide for applying techniques to real-world problems. Wiley-Interscience, New York

Yurtkuran A (2019) An improved electromagnetic field optimization for the global optimization problems. Comput Intel Neurosc 2019:6759106. https://doi.org/10.1155/2019/6759106

Zhu X, Lopes PE, Mackerell AD Jr (2012) Recent developments and applications of the CHARMM force fields. Wiley Interdiscip Rev Comput Mol Sci 2(1):167–185

Protein Folding, Dynamics and Aggregation at Single-Molecule Resolution

10

Ritobrita Chakraborty and Krishnananda Chattopadhyay

Abstract

The last couple of decades have witnessed a significant surge in the use of single-molecule fluorescence methods aimed at the in-depth understanding of how proteins fold to perform their biochemical functions. The detection of individual fluorophores enables the observation of a single protein molecule's conformational dynamic, in vivo localization within the heterogeneous cellular milieu, its fluorescence quenching upon intermolecular interactions, polarization response and fluorescence resonance energy transfer. This chapter has assembled a fundamental understanding of the single-molecule fluorescence techniques: Fluorescence Correlation Spectroscopy and Förster's Resonance Energy Transfer, and their use in case studies to illustrate how these techniques have enhanced our understanding of the mechanisms underlying protein folding and aggregation. Due to their high sensitivity and specificity, these techniques are becoming indispensable for the study of protein structure and conformational fluctuations.

Keywords

Single-molecule fluorescence · Fluorescence correlation spectroscopy · Energy landscape · Förster's resonance energy transfer · Alpha-Synuclein · Cytochrome c · Intestinal fatty acid-binding protein

R. Chakraborty · K. Chattopadhyay (✉)
Protein Folding and Dynamics Laboratory, Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata, India
e-mail: krish@iicb.res.in

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_10

10.1 Introduction

A comprehensive understanding of the mechanism of how a polypeptide chain folds into its three-dimensional functional form, or misfolds and assembles to form non-native aggregates, or how folded as well as natively unfolded/intrinsically disordered proteins (IDPs) transition between different conformational states (with the dynamics ranging between nanoseconds and seconds) necessitates a detailed exploration of the protein's energy landscape—which is typically inaccessible to classical ensemble time-averaging biophysical measurements. To circumvent the said bottleneck, single-molecule techniques that allow detailed examination of transient folding intermediates, heterogeneous sub-populations of misfolded or sparsely populated aggregated species (whose input within ensemble measurements may be lost) and rare events comprising short-lived structural species differing in their equally rapid internal dynamics and orientation have been harnessed in studying both ordered proteins and IDPs. Single-molecule techniques eliminate the necessity of synchronization in the presence of other biomolecules both in vitro and in vivo (Coelho et al. 2013) as well as are able to measure the kinetics between different stochastic states in a non-equilibrium system. The experimental results collected in the past few years highlight how achieving information on the smallest possible sub-population within an ensemble at the highest possible temporal resolution has enabled a scrupulous understanding of the folding, conformational dynamics and misfolding processes. In the case of IDPs (Uversky 2011; Sickmeier et al. 2007; Tompa 2002) in which the protein molecules can populate any given conformation within the continuous range between ordered and disordered structural forms (known as the disordered state ensemble, DSE, of the protein), (Bowler 2012) temporal resolution and duration of observation are crucial as the rapid interconversion dynamics of individual molecules might be lost to time averaging. Single-molecule study of folded proteins as well as IDPs has yielded previously unknown information about the folding landscapes, dynamics, multiple folding pathways, transient intermediates, unfolded ensembles and assisted folding of proteins.

Two major single-molecule spectroscopic techniques have enabled the dissection of the intricacies of the protein-folding process. The first is force spectroscopy, wherein mechanical unfolding of an ordered protein is achieved by towing it with laser tweezers, magnetic tweezers or an atomic force microscope. This technique provides insights about the energy landscape of the unfolding reaction coordinate under mechanical force (Onoa et al. 2003). The second technique is single-molecule fluorescence spectroscopy which allows the measurement of fluorescent signals from individual molecules, their analysis, and acquisition of information about the conformational dynamics involved in folding or aggregation/association of the single protein molecule. This chapter shall focus on the use of the latter technique on the in-depth understanding of protein folding, misfolding and association.

10.2 Single-Molecule Fluorescence as a Technique to Study Protein Folding, Dynamics and Aggregation

The 'fluorescence' phenomenon was first reported to occur in quinine solutions (Herschel 1851) and fluorspar (fluorite) (Brewster 1848). The process involves the excitation of an electron by a photon (timescale of 10^{-15} s) after which it returns to its ground state by dissipating the energy as light of longer wavelength/lower energy (timescale of 10^{-8} s). This relaxation is reliant on environmental factors, and properties of the emitted fluorescence (i.e., its wavelength, quantum efficiency, lifetime) can be harnessed as reporters of the changes in the environment surrounding the said fluorophore (Lakowicz 1999). The Jablonski diagram represents pictorially the energy levels involved in absorption and fluorescence (Fig. 10.1). The electronic ground state, first and second excited singlet states are designated as S_0, S_1 and S_2, respectively, while the thin horizontal lines represent vibrational levels. Transitions between the levels are shown as vertical arrows: straight ones depict radiative and curly ones show non-radiative transitions. At the ground S_0 state when an electron absorbs a photon, it promotes itself to an excited state, S_2, which is soon followed by a fast (timescale $\sim 10^{-12}$ s) non-radiative relaxation phenomenon accompanied by the release of heat, to the lowest vibrational level of S_1, in a process termed internal conversion (IC). From the S_1 stage, the excited molecule can either relax to the ground state by releasing a photon of higher wavelength (via fluorescence, timescale 10^{-8} s), or it can emit the excitation energy via internal conversion, without emitting a photon (IC, timescale $\sim 10^{-8}$ s), or the electrons can undergo a spin conversion to a triplet state (T_1), via intersystem crossing (ISC, timescale 10^{-8} to 10^{-3} s). This triplet state may subsequently decay to S_0 in a non-radiative way either via internal conversion or by emitting a photon via phosphorescence, which involves a longer timescale than fluorescence as it comprises a spin-forbidden transition.

In comparison to ensemble spectroscopic techniques which average over many particles, single-molecule fluorescence spectroscopy can resolve and quantify properties of individual protein conformers or their sub-populations which would

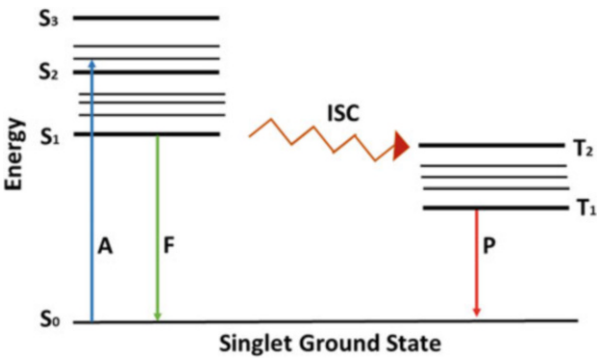


Fig. 10.1 The Jablonski diagram: A, F, and P represent photon absorption, fluorescence and phosphorescence emissions, respectively. Singlet and triplet states are indicated by S and T, respectively. ISC represents intersystem crossing (Figure prepared from Fig. 1 from Basak and Chattopadhyay, PCCP 2014; Reproduced by permission of the PCCP Owner Societies)

have been otherwise inaccessible within the energy landscape. The aim of this chapter is to focus on fluorescence spectroscopy, which due to its sensitivity and adaptability (Böhmer and Enderlein 2003), along with single-molecule Förster resonance energy transfer (smFRET) (Ha et al. 1996), enables investigation of inter-residue distances and conformational dynamics of individual protein molecules, thereby providing information on their folding, misfolding, association and function. Since the electrons reside in the excited state in S_1 comparatively longer (nanoseconds) than the excitation time frame (femtoseconds), the non-radiative phenomena (which includes FRET) compete with the radiative process of fluorescence. FRET is useful to study protein folding and association, and occurs when an excited electron of the donor fluorophore (D) transfers its energy to a second acceptor fluorophore (A), on condition that the emission spectrum of D and the absorption spectrum of A overlap to a suitable extent and that the distance between A and D is within of 2–10 nm. However, it is imperative to mention here that the large size of organic fluorescent dyes (used for tagging protein molecules) and their potential effects on the stability, folding and activity of the protein makes prior thermodynamic and kinetic characterization and comparison (with their unlabelled counterparts) of all the labelled species necessary.

10.2.1 Fluorescence Correlation Spectroscopy (FCS)

In principle, a process involving fluorescence intensity fluctuations can be quantified by correlation analysis (Elson 2011; Frieden et al. 2002). A fluorescence correlation spectroscopy experiment (Chattopadhyay et al. 2002, 2005) involves few protein molecules (at nanomolar concentration), freely diffusing through a confocal observation volume of a few femtolitres (Fig. 10.2). Any fluorescence fluctuation arising from the conformational dynamics, chemical reaction or association of the protein is then deciphered into an autocorrelation function which is fit to a suitable model with the intention of determining the translational diffusion time (τ_D) and diffusion coefficient (D) of the protein or its complex/aggregate. At equilibrium, although the average concentration of the protein species within the confocal volume remains constant over time, the fluctuations which take place due to the diffusion of the protein molecules or due to chemical kinetics can vary the fluorescence intensity. This change can be observed from the correlation function, only if the conformational fluctuation rate (τ_R) is much faster than diffusion ($\tau_R << \tau_D$). Notably, FCS measurements carried out using a low concentration of the protein with only a few molecules in the observation volume at a given point of time to avoid aggregation of predisposed systems. This is an improvement over traditional NMR and SAXS measurements using which the reported r_H values for α-synuclein (αSyn) have been shown to vary nearly twofold (Tashiro et al. 2008), possibly due to aggregation or oligomerization of the protein in solution at high concentrations ranging between 30 and 800 μM, which are required for such experiments.

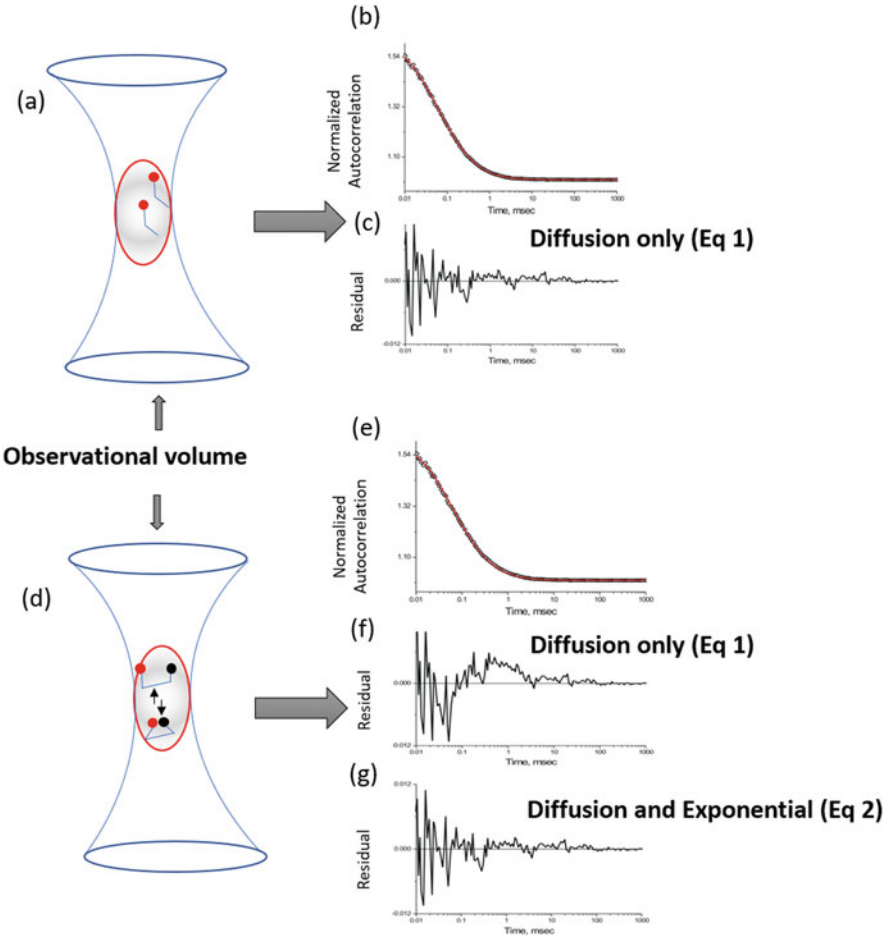


Fig. 10.2 A schematic diagram depicting a typical FCS experiment. (**a**) Fluorescence intensity fluctuations of the dye TMR are measured inside an observation volume of a few femtolitres. (**b**) The autocorrelation function, data are fit to a simple diffusion model without any exponential component (Eq. 10.1). (**c**) The goodness of the fit represented by the random residual distribution. (**d**) The confocal volume with diffusion with TMR-tagged haeme–protein cytochrome c molecules that show some conformational dynamics between an extended state, where the TMR (in red) and its quencher haeme (black) are distant from each other, and a compact 'dark' state (where TMR and haeme are close). (**e**) The correlation function obtained from this experiment could not be fit to the simple diffusion model, resulting in (**f**), a non-random residual distribution. (**g**) a fit using two components: diffusion and exponential (Eq. 10.2), which leads to random residual distribution. (Figure adapted from Haldar et al., J. Biol. Chem., 2010) (Haldar et al. 2010.)

The fluorescence intensity changes are quantified by auto-correlating the recorded intensity signal with respect to time, leading to the average number of fluorescent particles in the detection volume (N) and their average diffusion time (τ_D) through the volume. These fluctuations are mathematically expressed as the deviation of

instantaneous fluorescence intensity from the corresponding mean intensity over time, as follows:

$$\partial F(t) = F(t) - < F(t) >$$

and are correlated with time by the temporal autocorrelation function as follows:

$$G(\tau) = \frac{< \delta F(t) \cdot \delta F(t+\tau) >}{< F(t) >^2}$$

For one component diffusion, the function expressed in the form of parameters is as follows:

$$G(\tau) = 1 + \frac{1}{N} \cdot \left(\frac{1}{1 + \left(\frac{\tau}{\tau D}\right)} \right) \cdot \left(\frac{1}{\sqrt{1 + S^2 \left(\frac{\tau}{\tau D}\right)}} \right) \tag{10.1}$$

where τ_D is the average diffusion time of the fluorescently labelled protein, and S defines the ratio of beam radius r_0 to beam height z_0 of the instrument. τ_D is related to the diffusion coefficient (D) as per the equation:

$$\tau_D = \omega^2 / 4D$$

where ω is the size of the observation volume.

Additionally, the value of the hydrodynamic radius (r_H) of the protein molecule/complex/aggregate can be obtained from D using the Stokes−Einstein formalism:

$$D = \mathrm{kT}/6\pi\eta r_H$$

The correlation function for a single component system with diffusion time τ_D with an associated event of either a chemical reaction or conformational change (A→B and B→A) with a relaxation time constant of τ_R can be described by Eq. 10.2:

$$G(\tau) = \frac{\left(1 - F + F_{\exp}\left(-\frac{\tau}{\tau_R}\right)\right)}{N(1-F)} \cdot \left[(1 + \tau/\tau_D) \cdot \left(1 + \frac{\tau}{\omega^2 \tau_D} \right)^{0.5} \right]^{-1} \tag{10.2}$$

The above equation assumes that one of the states (A or B) involving the conformational change event is non-fluorescent (dark), and F (the amplitude of relaxation time, τ_R) is the average fraction of this non-fluorescent conformer. Equation 10.2 also assumes that A and B have the same diffusion time. The value of τ_R can only be measured if the rate of a chemical reaction or the conformational change event occurs significantly faster than the diffusion time of the protein molecule ($\tau_R \ll \tau_D$). By monitoring the changes in τ_D or r_H, with respect to

environmental conditions, intermediates in protein folding, their oligomerization (Rajagopalan et al. 2010) or aggregation can be observed.

10.2.2 Förster Resonance Energy Transfer (FRET)

SmFRET is based on the transfer of energy from an excited-state donor (D) to an acceptor chromophore (A), separated by no more than 2–10 nm (Schuler and Hofmann 2013), via non-radiative dipole–dipole coupling. Once excited by the absorption of light, the donor undergoes various de-excitation pathways apart from FRET, such as fluorescence (rate constant given by k_f) as well as other non-radiative decays (total rate constant represented ask$_{nr}$) so that the efficiency of energy transfer by FRET is expressed as $E_{FRET} = k_{FRET}/(k_f + k_{FRET} + k_{nr})$.

This efficiency is calculable in terms of the experimental observables such as fluorescence lifetimes and fluorescence intensities of the donor molecule using the following equations:

$$E_{FRET} = 1 - \left(\frac{\tau_{D'}}{\tau_D}\right) \text{and } E_{FRET} = 1 - \left(\frac{F_{D'}}{F_D}\right) \text{respectively,}$$

where $\tau_{D'}$and τ_D are the donor fluorescence lifetimes in the presence and absence of an acceptor, while $F_{D'}$ and F_D are the donor fluorescence intensities in the presence and absence of an acceptor.

In case of smFRET, the transfer efficiency is expressed as the number of photons collected from the acceptor and donor (n_A and n_D, respectively):

$$E_{FRET} = n_A/(n_A + {}^{\gamma}_{D}n)$$

where γ is the correction factor between the quantum yields of the fluorophore and the detection efficiency of the instrument.

From the FRET efficiency thus calculated, the distance between donor and acceptor can be obtained, by using the following equation:

$$E_{FRET} = 1/\left[1 + \left(\frac{r}{R_0}\right)^6\right] \tag{10.3}$$

where R_0 is the Forster distance at which the energy transfer efficiency is 50%, theoretically calculable for a particular system of donor and acceptor.

By labelling two different loci on the same protein with the donor and acceptor fluorophore, the measurement of '*r*' (or rather its distribution over the ensemble of conformers) is possible, which provides insight into the relative positioning of the two sites within the protein, thus shedding light onto its structure. The structural dynamics of the protein can also be studied from a variation of the FRET efficiency over time. The same time-dependent study carried out on two interacting proteins by

labelling one with the donor and the other with the acceptor, dynamics of the interaction can be studied.

In an IDP ensemble, FRET experiments can separate signals from various sub-populations of conformers, thus enabling the measurement of parameters of each sub-population (Widengren et al. 2006). For example, in a confocal experiment with pulsed excitation and four detection channels, the emission wavelength range (either a donor or an acceptor photon), the polarization and the time of emission relative to the excitation pulse are available for every detected photon. Therefore, parameters like transfer efficiency, donor and acceptor fluorescence lifetimes, fluorescence anisotropy required for accurate distance calculation can be obtained for each burst of photons from a single molecule (Schuler and Hofmann 2013; Sisamakis et al. 2010). Bursts from individual sub-populations can be grouped to obtain their fluorescence decays in a detailed manner without interference from other molecules/sub-populations. Therefore, apart from providing additional information, observational difficulties such as limited rotational mobility of chromophores (Hillger et al. 2008) or fluorescence quenching due to the complex formation (Sisamakis et al. 2010) can also be identified for analysis.

10.3 Understanding the Particulars of Protein Folding at Single-Molecule Detail

Thorough insights into protein folding entail understanding closely related questions about the conformational thermodynamics determining a protein's native three-dimensional structure from its linear amino acid sequence, as well as its surrounding environment. Unanswered queries pertaining to the kinetics, timescales, mechanisms (whether assisted by interactions with ligands, molecular chaperones, protein partners or small molecule osmolytes) and energy landscapes (shape and contour: rugged versus smooth) of folding, as well as predicting the protein's native structure also require understanding (Dill et al. 2008). Further questions comprise how the presence of thermodynamic states and intermediates, and which aspects of their structural properties govern protein folding, as well as whether these intermediates lead to the folded protein (on-pathway intermediates) or kinetically trapped intermediates (off-pathway). Additionally, with respect to IDPs, the structures (random coil versus residual structure), dynamics or fluctuations (as well as the extent of fluctuations) in each thermodynamic state of the denatured state ensemble need to be determined. Additional questions include the characteristics of the energy barriers that exist in the folding transitions as well as the kind of interactions, whether enthalpic or entropic, involved in protein folding. The answers to these questions are much needed as they shall shed insight into de novo protein design and enable strategies for combating protein misfolding diseases (Basak and Chattopadhyay 2014).

10.3.1 FCS Investigations of the Protein Folding Landscape

10.3.1.1 Unfolded State Conformers and Folding Intermediates of a Non-amyloid-Forming β-Sheet Protein

The intestinal fatty acid-binding protein (IFABP, 15 kDa) is a classic example for understanding the folding behaviour of β-sheet proteins due to its small size, stability, lack of proline and cysteine residues and monomeric nature. It consists of two β-sheets, each comprising five β-strands, along with a short helical region. The protein binds ligands such as fatty acids, bile salts and retinoids within its central cavity. Equilibrium unfolding transitions monitored by ensemble methods like steady state fluorescence and CD measurements showed a typical two-state unfolding. However, unfolding experiments using FCS in the presence of decreasing pH show a compact intermediate portrayed by a small upsurge in the diffusion coefficient around pH 3.5, before the protein completely unfolds at lower pH values (Chattopadhyay et al. 2002). The acid unfolded state of the protein at a pH lower than 3.5 shows a high diffusion coefficient but is not a random coil (Chattopadhyay et al. 2002), while the intermediate at pH 3.5 has a strong helix-forming inclination and is prone to aggregation (Sarkar-Banerjee et al. 2016). This helix formation occurs due to the interaction of amino acids situated around the hydrophobic core residues. Further structural reorganization triggered by distant contacts, steric constraints of the hydrophobic side chains and unfavourable entropic costs leads to the reorganization of the early contacts resulting in the formation of the functional form.

The existence of the extended and compact conformers of the unfolded state of IFABP has been studied previously using X-ray scattering, FRET and FCS (Ziv and Haran 2009). Sarkar et al., using FCS supplemented with mutagenesis of key residues (Sarkar and Chattopadhyay 2014), demonstrated that the early collapse during the folding process of IFABP occurs as a result of hydrogen bonding (and is not entirely a hydrophobic collapse) and is dependent on the protein sequence. The unfolded state of IFABP contains some transient residual structure around the D-E turn, indicating that these residues may initiate the folding following which formation of medium- and long-range contacts develops the overall secondary structure (Ropson et al. 2006; Hodsdon and Frieden 2001). Similar studies demonstrating the role of the backbone hydrogen bond formation during early collapse have been observed elsewhere (Bowler 2012; Bolen and Rose 2008). Any change in the form of an increase in the hydrophobicity in the D-E turn or in the loop region could lead to the formation of misfolded contacts that could then potentially result in aggregation. These observations from FCS studies reveal the relatedness between an early hydrophobic collapse, secondary structure formation, and misfolding-aggregation all of which are crucial factors in determining the protein folding process. The study of all these events at a single-molecule resolution enables a comprehensive understanding of three-dimensional protein folding.

10.3.1.2 Unfolded State Conformation and Early Folding Kinetics of the α-Helical Ordered Protein Cytochrome *c*

Furthermore, FCS has also unravelled the presence of an intermediate formed within the unfolding pathway of bovine serum albumin (BSA), a predominantly helical protein (Ghosh et al. 2009), whose unfolding transition monitored by ensemble measurements (far-UV CD and tryptophan fluorescence) could only be fit to two-state models. This intermediate state is formed at a low urea concentration before the complete unfolding of the secondary structure occurs at a higher concentration of urea (Ghosh et al. 2009). Interestingly, this intermediate is a precursor of aggregation and does not form in the presence of arginine, which is an inhibitor of aggregation (Arakawa et al. 2007a, b) and protein self-association (Ropson and Frieden 1992; Basak and Chattopadhyay 2013; Chattopadhyay et al. 2002).

Arginine, among other osmolytes, has been shown using FCS, to cause the compaction of unfolded conformations of proteins, as it is a potent stabilizer for improving the yield of proteins during refolding (Haldar et al. 2010). An understanding of the conformation and dynamics of unfolded and early-stage intermediate states of a protein enables deciphering of its mechanism of aggregation which in turn can aid in the design of potent inhibitor molecules. The helical protein cytochrome *c* from *Saccharomyces cerevisiae* (y-cytc) has been studied at single molecular resolution using FCS with ease as the presence of the haeme cofactor coupled with an extrinsically attached dye constitutes a fluorescence resonance energy transfer pair, which can indicate the dynamics of individual protein molecules. Furthermore, the role of the protein stabilizers such as sucrose, sodium chloride, proline and arginine on the conformation of the unfolded state and the kinetics early-stage folding of y-cytc was studied. The results showed that arginine stabilizes a condensed form of the unfolded protein. During the urea-induced unfolding of TMR-cytc in the presence and absence of arginine, the authors were able to show that a high arginine concentration caused the formation of a partially folded intermediate. Such a phenomenon is not observed for the other stabilizers under study, namely sucrose, sodium chloride and proline. The hydrodynamic radius (rH) was found to increase in a two-state form when y-cytc is unfolded by urea. Moreover, arginine and other stabilizers led to the formation of a comparatively structured conformation of the unfolded state (with an rH of 29 Å). An extended conformer with an rH of 40 Å is formed in the absence of such stabilizers.

In order to study the folding and function of cytochrome c within living cells, the effect of the crowded cellular milieu was replicated using synthetic crowding media after which the protein was studied using FCS. Under native conditions in aqueous buffer, y-cytc is populated by an expanded conformer in equilibrium with a compact conformer. The crowding medium (ficoll/dextran) could modify this equilibrium between the compact and expanded forms, causing an increase in the formation of the compact conformer population resulting in the decrease in peroxidase activity of y-cytc (as compaction of the protein causes a decrease in haeme-surrounding solvent exposure, which is the main causative agent of peroxidase activity). Interestingly, urea-induced protein stability measurements show that compaction due to crowding

agents destabilizes y-cytc due to electrostatic repulsions between similarly charged clusters located on the protein (Paul et al. 2016).

10.4 Understanding the Workings of Protein Misfolding and Aggregation at Single-Molecule Resolution

The aggregation of proteins is believed to be dependent on the sequence of a protein and therefore it competes with native folding (Jaenicke 1995). Self-association of monomers which either have their native conformations or are partially denatured or misfolded leads to oligomerization. Such oligomers are of various sizes, shapes and have varying extents of hydrophobic exposure or cytotoxicity. These, in turn, append to form larger aggregates, which may be either amorphous or ordered such as amyloid fibrils (Eichner and Radford 2011). In order to understand the technicalities of the aggregation process, (1) identification and characterization of the different structural species formed during the various stages of aggregation, (2) elucidation of the various pathways through which these different species aggregate through the formation of stable and metastable intermediates (3) and an understanding of the effects of protein modifications (Ellis et al. 2012) and mutations (Baskakov et al. 2005), co-factors, additives and co-solvents on the aggregation pathways are essential.

The structural features of the species within an aggregation pathway influence properties like protease resistance, solubility, protein–protein as well as membrane association, all of which may influence their membrane pore formation and cytotoxicity (Kagan 2012; Olzscha et al. 2011). Other characteristics of the oligomers governed by their structure include the rate of their dissociation from or association into higher-order aggregates, their localization within a cell (Winkler et al. 2010), and ability to be trafficked between cells (Goedert et al. 2010), which impacts the prion-like disease propagation observed in many misfolding diseases (Aguzzi and Rajendran 2009; Grad et al. 2011; Stohr et al. 2012). Mapping the pathways that link the aggregation states enables identification of the rate-limiting steps that are crucial for understanding the aggregation mechanism of the protein (Roberts 2007). Due to the heterogeneity and intricacy of the aggregation landscape, different species from different stages/competing pathways of aggregation may populate the landscape simultaneously (Kodali and Wetzel 2007). Consequently, the properties of an aggregating protein are determined by the ensemble of its aggregated forms whose composition varies temporally. Importantly, in aggregated systems, if the relative concentration of the cytotoxic aggregate species is low, ensemble biophysical methods cannot detect their presence in the aggregate ensemble.

The typical practices used to observe protein folding when used for analysing the process of aggregation (Nilsson 2004; Dobson 2004) cause an averaging of the collection of the different states present in the sample at a given time. Such averaging challenges the interpretation because key conformations that form rarely or briefly may be undetected. Single-molecule approaches monitor molecules individually such that the effects of ensemble averaging are circumvented. Both

sub-populations and rare events/conformers can be detected directly, as well as solitary molecules can be pursued temporally, due to which the whole network of intermediate states and their conversion time points can be mapped. In the following section, we shall focus on the use of single-molecule fluorescence spectroscopy to delve deeper into the processes and intermediates formed during aggregation.

10.4.1 Single-Molecule Investigations of the Protein Misfolding-Aggregation Landscape

10.4.1.1 Early-Stage Intra-Molecular Collapse and Subsequent Intermolecular Binding and Folding or Aggregation of the IDP αSyn

The monomers of the IDP αSyn misfold and coalesce to form ordered cross β-sheet amyloid fibrils which assemble as cytosolic plaques within the dopaminergic neurons in the mid-brain of persons suffering from the neurodegenerative disorder Parkinson's disease (Chakraborty et al. 2019; Spillantini et al. 1997). The disordered state ensemble (DSE) (Bowler 2012) of αSyn undergoes several structural reorganizations while folding through the formation of an early-stage collapsed state (Morar et al. 2001) followed by either the folded form or misfolded aggregates, all of which depend upon solvent conditions. Single-molecule fluorescence spectroscopy has been employed for probing such self-association and intermolecular interactions of αSyn. Due to its role in neurodegenerative disorders, αSyn and its interactions with lipids and membrane mimetics (Ferreon et al. 2010; Trexler and Rhoades 2013; Drescher et al. 2012) are being widely studied using single-molecule techniques. Using a combination of FRET and FCS (Basak et al. 2015), the effects of the membrane-mimetic sodium dodecyl sulphate (SDS) on the structural transitions between an intra-chain collapsed form, which occurs at a very low concentration (lesser than the critical micellar concentration, CMC) of SDS, and inter-chain aggregate at a concentration close to the CMC of SDS were studied at single-molecule resolution (Fig. 10.3). At a concentration above the CMC (of ~1 mM), αSyn was shown to bind to the SDS micelles in a broken helix conformation, while at higher concentrations of SDS, characterized by cylindrical micelles or extended bilayers of SDS, an extended conformation of αSyn was formed, which resembled the form of αSyn bound to vesicles (Ferreon et al. 2009; Veldhuis et al. 2009). Thus, it is the concentration of SDS and its state of assembly that dominates the conformational distribution of the IDP αSyn.

The extent of early-stage compaction/collapse has been found to correlate with later-stage aggregation. In a related report, it was studied using FRET and FCS that the conformers of αSyn switch between a broken/kinked helix (i.e. a helix turn helix structure) and an extended helix depending upon the concentration as well as curvature of the binding membrane (micelles versus bilayers) (Ferreon et al. 2009). It was also observed that when αSyn binds to lipid vesicles containing a lesser percentage of negatively charged lipids (this composition is similar to

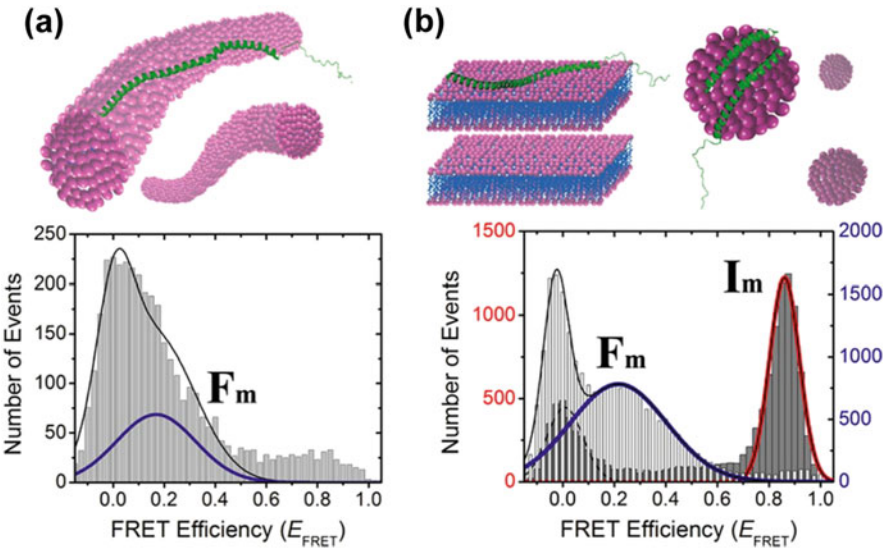


Fig. 10.3 αSyn folding induced by interaction with the lipid membrane-mimic SDS. (**a**) In the presence of a very high concentration of SDS (450 mM), low FRET efficiency (EFRET) is observed, as an extended helical structure bound to cylindrical micelles is formed, represented as Fm. (**b**) Subsequent addition of co-surfactant hexanol transforms the spherical micelles (concentration 50 mM) into flat bilayers, which induces a protein conformational switch from the high-EFRET spherical micelle-bound Im form (red curve) to the low-EFRET bilayer-bound Fm species (blue curve; overlaid histogram; thin lines: Gaussian fit to data; bold lines: fitted peaks for indicated conformations). Figure prepared from figure 3 of Ferreon et al. Proc. Natl. Acad. Sci. USA, 2009 (Ferreon et al. 2009)

biological membranes), it forms an extended helix, implying that the two-folded structures (broken and extended) are tuneable by differing the membrane properties.

The inherent heterogeneity of folding-misfolding-aggregation landscape of αSyn and the short-lived nature of the early-stage intermediates/oligomers complicate the development of a successful therapeutic intervention. The role of co-solvents on the compaction-aggregation of αSyn was investigated recently by Ghosh et al. (Fig. 10.4). An inhibitor of aggregation, arginine, was shown to delay the late events of amyloid fibril formation, by binding to the protein and forming a nearly compact state (Ghosh et al. 2018). In contrast, glutamate, which is a facilitator of aggregation, was found to be completely excluded from the surface of the protein and initiate and hasten the process of aggregation. These opposing effects of the inhibitor and facilitator were additive, and together they were shown to maintain a ratio by which they could cancel out each other's effect at the different stages of αSyn aggregation.

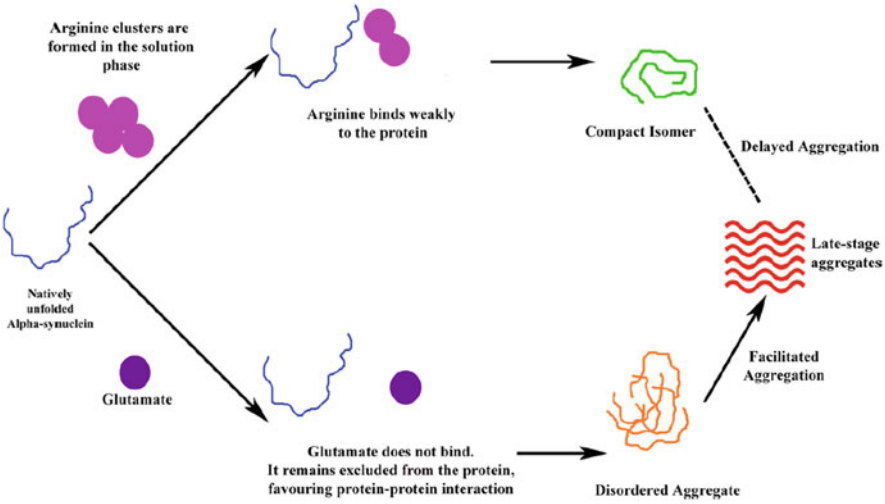


Fig. 10.4 Mechanistic outline of the effect of arginine and glutamate on the αSyn aggregation landscape (Figure created from figure 9 from Ghosh et al., Sci. Rep., 2018) (Ghosh et al. 2018)

10.4.1.2 Use of smFRET to Characterize the Early-Stage Oligomer Formation of αSyn

Cremades et al. (2012) established a single-molecule fluorescence technique that enables the exhaustive characterization of the oligomerization process of αSyn (Fig. 10.5). In this method, a mixture of αSyn labelled with either AlexaFluor 488 or AlexaFluor 594 is incubated under aggregation-inducing conditions, after which at regular time points, aliquots are diluted for detection. The oligomers comprise both AlexaFluor 488 and AlexaFluor 594 labelled monomers and give rise to a simultaneous burst of fluorescence intensity in both AlexaFluor 488 and AlexaFluor 594 channels. As the two dyes are a FRET pair, the oligomer fluorescence intensities can be used to estimate oligomer size in addition to FRET efficiency. The group has also supplemented fast-flow microfluidic techniques to the single-molecule fluorescence technique to increase the rate of data acquisition. They identified an important slow-conversion step from the initially formed oligomers (type-A), which possessed low FRET efficiency, high sensitivity to proteinase K digestion and low toxicity into the more compact oligomers (type-B) that displayed higher FRET efficiency values, a higher resistance to proteolysis, presence of a beta-sheet folding core and high cytotoxicity.

10.4.1.3 Insights into the Function and Aggregation of the IDP Tau Obtained from FCS and FRET Studies

Tau, a highly dynamic microtubule-associated IDP, possesses significant functions in regulating the dynamic instability of microtubules within the neuronal axons. Neurodegenerative diseases such as Alzheimer's disease caused by tau are characterized by the loss of its native function as well as its aggregation (Lee et al.

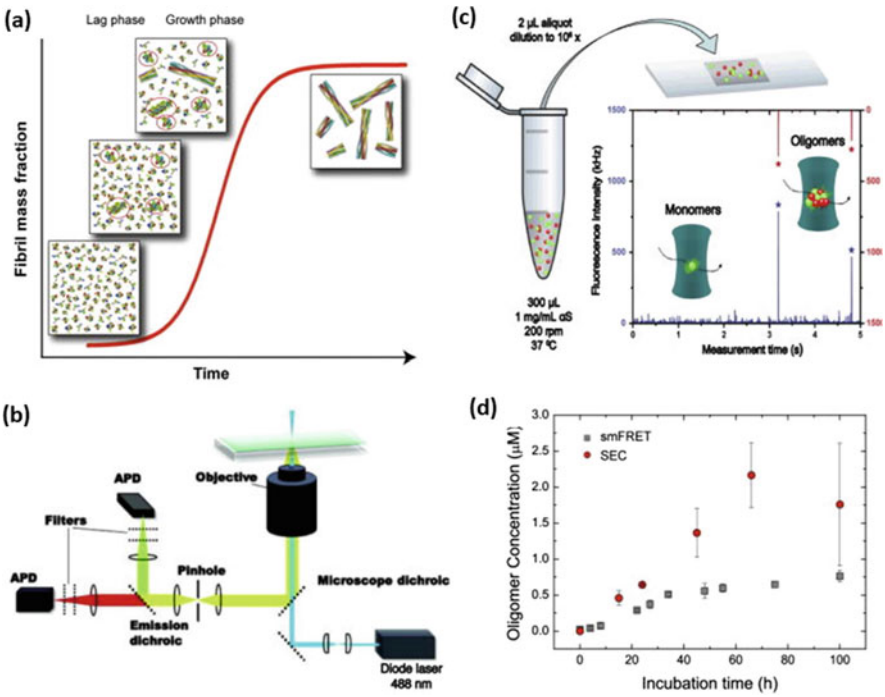


Fig. 10.5 (**a**) Kinetics of αSyn fibrillation; oligomeric species highlighted with red circles. (**b**) Schematic representation of the smFRET setup. (**c**) Schematic description of the experimental protocol: fluorescent bursts in both channels indicate the presence of FRET-positive oligomeric species (asterisks). Non coincident bursts are attributed to monomers and are much less bright. (**d**) Comparison between the kinetics of oligomer formation under bulk conditions obtained by quantitative SEC analysis (red circles) and smFRET experiments (the concentrations from single-molecule experiments have been extrapolated to bulk conditions (grey squares)). Figure reproduced from Figure 1 of Cremades et al., Cell, 2012 (Cremades et al. 2012)

2001). An early ensemble FRET study (Jeganathan et al. 2006) reported Tau to form a compacted folded structure with long-range contacts connecting its N- and C-termini, as well as to contain a function-related microtubule binding domain (MTBD) even in its globally disordered native state, from which the authors surmised a 'paper-clip'-like structure for tau. Recent investigations (Elbaum-Garfinkle and Rhoades 2012) have since recognized individual domains within tau using smFRET. In an important study, the conformational changes caused due to heparin-induced aggregation and tubulin-induced tau function have been studied using FCS and FRET (Melo et al. 2017). These techniques have been of specific use as the critical initial stages of the processes of tau aggregation and folding are masked by the averaging effect of ensemble measurements. Tau binding to heparin displayed a cooperative shift to a distinct conformation, characterized by an overall surge in protein dimension, along with the microtubule binding domain (MTBD),

which forms the functional core of the fibrils, exhibiting a two-state transition to a compact state.

10.5 Future Perspectives

Research using single-molecule techniques is harnessed to provide an understanding of the protein folding problem, as well as comprehend in detail, misfolding and aggregation specifics, which is imperative for deciphering their link to human disease, and developing therapeutics against them. From these studies, the generic principle of protein folding that occurs via an initial hydrophobic/backbone-based collapse followed by secondary structure formation has emerged. It is now believed that the fluctuation of a protein molecule between folding and misfolding/aggregation is dependent on its sequence, as well as solution conditions, presence of co-factors and cellular stress. Although techniques such as microscopy (including cryo-electron microscopy) can provide a visualization of the different structural states of a protein, single-molecule techniques have an advantage as they can illuminate the connecting pathways between these states. The development of correlations between the biophysical observations of such conformers and their disease-causing characteristics (including membrane ion-channel formation propensity and neuronal death) could be a future undertaking. Single molecule of protein folding/aggregation landscapes can also reveal a common phenomenon by which oligomers are formed and how they inflict toxicity on cells. The use of such knowledge can then be utilized to rationally strategize therapies against these disease-inducing pharmacological targets.

Researchers are combining microfluidics with single-molecule FRET measurements (Lemke et al. 2009; Pfeil et al. 2009) which has resulted in improved detection, a reduced amount of photo bleaching and increased fluorophore brightness. Superior detection of ultra-fast biomolecular dynamics in the μs timescale is much needed to study proteins, which fold near the folding speed limit (Bryngelson et al. 1995). Rapid and high-resolution analyses of folding landscapes can enable a much-needed exploration of the transition region of folding. Development of methods for performing three- and multi-colour FRET measurements (Gambin and Deniz 2010; Sisamakis et al. 2010) permits the understanding of correlated conformational changes in large multi-domain protein complexes and protein–protein interactions.

Another important future area of protein folding research comprises studying the in-cell folding and aggregation behaviour of proteins, which is otherwise difficult to probe by two-component in vitro systems. Development in the studies of co-translational folding has provided a technique to explore the workings of in-cell protein folding. Complex characteristics of the kinetics and thermodynamics of IDPs that fold upon binding with model membranes are also being unravelled by single-molecule techniques.

10.6 Conclusions

Biomolecules are characterized by multiple metastable states that are involved in complex inter-conversions within an active milieu. In ensemble fluorescence measurements, individual properties of such inter-converting molecules remain concealed due to collective averaging. Single-molecule techniques, on the other hand, allow the detection of emission from individual molecules (rather than statistical averaging) so as to identify hidden heterogeneity within a system. Thus, the characteristics of rare intermediate conformers, differences in reaction pathways, accurate distance measurements between two chromophores within the same molecule, etc., become tangible when studied by single-molecule experiments. Such experiments utilize a minimal detection volume to record fluorescence bursts of discrete molecules at very diluted conditions within the femtolitre observation volume, thereby increasing the signal-to-noise ratio. Developments over the years, in experimental and computational techniques, include improvements in sensitivity and speed of detectors, stability and efficacy of illumination sources and probes which aided the established of single-molecule fluorescence as a mainstream biophysical tool to detect molecules or manipulate them. A wide range of properties of the emission from the fluorophore are now being obtained, such as polarization, spectrum, degree of energy transfer, and spatial position. The time dependence of each parameter harvests information about the excited-state lifetimes, photochemistry, local environmental fluctuations and biological activity of the protein. This chapter aims to provide the reader with a basic understanding of the principles governing single-molecule fluorescence-based techniques, and applications aimed to answer biochemical questions underlying protein folding, dynamics, and function/disease, using both α-helical and β-sheet-rich proteins, as well as aggregation-prone IDPs as examples, thus appealing to the reader to further pursue the literature references linked to the examples discussed.

References

Aguzzi A, Rajendran L (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. Neuron 64(6):783–790

Arakawa T, Ejima D, Tsumoto K, Obeyama N, Tanaka Y, Kita Y, Timasheff SN (2007a) Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. Biophys Chem 127(1–2):1–8

Arakawa T, Tsumoto K, Nagase K, Ejima D (2007b) The effects of arginine on protein binding and elution in hydrophobic interaction and ion-exchange chromatography. Protein Expr Purif 54 (1):110–116

Basak S, Chattopadhyay K (2013) Fluorescence correlation spectroscopy study on the effects of the shape and size of a protein on its diffusion inside a crowded environment. Langmuir 29 (47):14709–14717

Basak S, Chattopadhyay K (2014) Studies of protein folding and dynamics using single molecule fluorescence spectroscopy. Phys Chem Chem Phys 16(23):11139–11149

Basak S, Prasad GV, Varkey J, Chattopadhyay K (2015) Early sodium dodecyl sulfate induced collapse of alpha-synuclein correlates with its amyloid formation. ACS Chem Neurosci 6 (2):239–246

Baskakov I, Disterer P, Breydo L, Shaw M, Gill A, James W, Tahiri-Alaoui A (2005) The presence of valine at residue 129 in human prion protein accelerates amyloid formation. FEBS Lett 579 (12):2589–2596

Böhmer M, Enderlein J (2003) Fluorescence spectroscopy of single molecules under ambient conditions: methodology and technology. ChemPhysChem 4(8):792–808

Bolen DW, Rose GD (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu Rev Biochem 77:339–362

Bowler BE (2012) Residual structure in unfolded proteins. Curr Opin Struct Biol 22(1):4–13

Brewster D (1848) On the decomposition and dispersion of light within solid and fluid bodies. Lond Edinb Dublin Philos Mag J Sci 32(217):401–412

Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21(3):167–195

Chakraborty R, Dey S, Paul SS, Sil P, Sengupta J, Chattopadhyay K (2019) Heme minimizes Parkinson's disease-associated toxicity by inducing a conformational distortion in the oligomers of alpha-Synuclein. bioRxiv 629238

Chattopadhyay K, Saffarian S, Elson EL, Frieden C (2002) Measurement of microsecond dynamic motion in the intestinal fatty acid binding protein by using fluorescence correlation spectroscopy. Proc Natl Acad Sci U S A 99(22):14171–14176

Chattopadhyay K, Elson EL, Frieden C (2005) The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. Proc Natl Acad Sci U S A 102 (7):2385–2389

Coelho M, Maghelli N, Tolić-Nørrelykke IM (2013) Single-molecule imaging in vivo: the dancing building blocks of the cell. Integr Biol 5(5):748–758

Cremades N, Cohen SIA, Deas E, Abramov AY, Chen AY, Orte A, Sandal M, Clarke RW, Dunne P, Aprile FA, Bertoncini CW, Wood NW, Knowles TPJ, Dobson CM, Klenerman D (2012) Direct observation of the interconversion of normal and toxic forms of α-synuclein. Cell 149(5):1048–1059

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

Dobson CM (2004) Experimental investigation of protein folding and misfolding. Methods 34 (1):4–14

Drescher M, Huber M, Subramaniam V (2012) Hunting the chameleon: structural conformations of the intrinsically disordered protein alpha-synuclein. Chembiochem 13(6):761–768

Eichner T, Radford SE (2011) A diversity of assembly mechanisms of a generic amyloid fold. Mol Cell 43(1):8–18

Elbaum-Garfinkle S, Rhoades E (2012) Identification of an aggregation-prone structure of tau. J Am Chem Soc 134(40):16607–16613

Ellis CR, Maiti B, Noid WG (2012) Specific and nonspecific effects of glycosylation. J Am Chem Soc 134(19):8184–8193

Elson EL (2011) Fluorescence correlation spectroscopy: past, present, future. Biophys J 101 (12):2855–2870

Ferreon AC, Gambin Y, Lemke EA, Deniz AA (2009) Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. Proc Natl Acad Sci U S A 106(14):5645–5650

Ferreon AC, Moran CR, Gambin Y, Deniz AA (2010) Single-molecule fluorescence studies of intrinsically disordered proteins. Methods Enzymol 472:179–204

Frieden C, Chattopadhyay K, Elson EL (2002) What fluorescence correlation spectroscopy can tell us about unfolded proteins. Adv Protein Chem 62:91–109

Gambin Y, Deniz AA (2010) Multicolor single-molecule FRET to explore protein folding and binding. Mol BioSyst 6(9):1540–1547

Ghosh R, Sharma S, Chattopadhyay K (2009) Effect of arginine on protein aggregation studied by fluorescence correlation spectroscopy and other biophysical methods. Biochemistry 48 (5):1135–1143

Ghosh S, Kundu A, Chattopadhyay K (2018) Small molecules attenuate the interplay between conformational fluctuations, early oligomerization and amyloidosis of alpha synuclein. Sci Rep 8(1):5481

Goedert M, Clavaguera F, Tolnay M (2010) The propagation of prion-like protein inclusions in neurodegenerative diseases. Trends Neurosci 33(7):317–325

Grad LI, Guest WC, Yanai A, Pokrishevsky E, O'Neill MA, Gibbs E, Semenchenko V, Yousefi M, Wishart DS, Plotkin SS, Cashman NR (2011) Intermolecular transmission of superoxide dismutase 1 misfolding in living cells. Proc Natl Acad Sci U S A 108(39):16398–16403

Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, Weiss S (1996) Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. Proc Natl Acad Sci U S A 93(13):6264–6268

Haldar S, Mitra S, Chattopadhyay K (2010) Role of protein stabilizers on the conformation of the unfolded state of cytochrome c and its early folding kinetics: investigation at single molecular resolution. J Biol Chem 285(33):25314–25323

Herschel JFW (1851) On a case of superficial colour presented by a homogeneous liquid internally colourless. Proc R Soc Lond 5:547–547

Hillger F, Hanni D, Nettels D, Geister S, Grandin M, Textor M, Schuler B (2008) Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. Angew Chem Int Ed Engl 47(33):6184–6188

Hodsdon ME, Frieden C (2001) Intestinal fatty acid binding protein: the folding mechanism as determined by NMR studies. Biochemistry 40(3):732–742

Jaenicke R (1995) Folding and association versus misfolding and aggregation of proteins. Philos Trans R Soc Lond Ser B Biol Sci 348(1323):97–105

Jeganathan S, von Bergen M, Brutlach H, Steinhoff HJ, Mandelkow E (2006) Global hairpin folding of tau in solution. Biochemistry 45(7):2283–2293

Kagan BL (2012) Membrane pores in the pathogenesis of neurodegenerative disease. Prog Mol Biol Transl Sci 107:295–325

Kodali R, Wetzel R (2007) Polymorphism in the intermediates and products of amyloid assembly. Curr Opin Struct Biol 17(1):48–57

Lakowicz JR (1999) Principles of fluorescence spectroscopy, 2nd edn. Kluwer Academic/Plenum, New York

Lee VM, Goedert M, Trojanowski JQ (2001) Neurodegenerative tauopathies. Annu Rev Neurosci 24:1121–1159

Lemke EA, Gambin Y, Vandelinder V, Brustad EM, Liu HW, Schultz PG, Groisman A, Deniz AA (2009) Microfluidic device for single-molecule experiments with enhanced photostability. J Am Chem Soc 131(38):13610–13612

Melo AM, Elbaum-Garfinkle S, Rhoades E (2017) Insights into tau function and dysfunction through single-molecule fluorescence. Methods Cell Biol 141:27–44

Morar AS, Olteanu A, Young GB, Pielak GJ (2001) Solvent-induced collapse of alpha-synuclein and acid-denatured cytochrome c. Protein Sci 10(11):2195–2199

Nilsson MR (2004) Techniques to study amyloid fibril formation in vitro. Methods 34(1):151–160

Olzscha H, Schermann SM, Woerner AC, Pinkert S, Hecht MH, Tartaglia GG, Vendruscolo M, Hayer-Hartl M, Hartl FU, Vabulas RM (2011) Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. Cell 144(1):67–78

Onoa B, Dumont S, Liphardt J, Smith SB, Tinoco I Jr, Bustamante C (2003) Identifying kinetic barriers to mechanical unfolding of the T. thermophila ribozyme. Science 299 (5614):1892–1895

Paul SS, Sil P, Chakraborty R, Haldar S, Chattopadhyay K (2016) Molecular crowding affects the conformational fluctuations, peroxidase activity, and folding landscape of yeast cytochrome c. Biochemistry 55(16):2332–2343

Pfeil SH, Wickersham CE, Hoffmann A, Lipman EA (2009) A microfluidic mixing system for single-molecule measurements. Rev Sci Instrum 80(5):3125643

Rajagopalan S, Huang F, Fersht AR (2010) Single-molecule characterization of oligomerization kinetics and equilibria of the tumor suppressor p53. Nucleic Acids Res 39(6):2294–2303

Roberts CJ (2007) Non-native protein aggregation kinetics. Biotechnol Bioeng 98(5):927–938

Ropson IJ, Frieden C (1992) Dynamic NMR spectral analysis and protein folding: identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by 19F NMR. Proc Natl Acad Sci U S A 89(15):7222–7226

Ropson IJ, Boyer JA, Dalessio PM (2006) A residual structure in unfolded intestinal fatty acid binding protein consists of amino acids that are neighbors in the native state. Biochemistry 45 (8):2608–2617

Sarkar S, Chattopadhyay K (2014) Studies of early events of folding of a predominately beta-sheet protein using fluorescence correlation spectroscopy and other biophysical methods. Biochemistry 53(9):1393–1402

Sarkar-Banerjee S, Chowdhury S, Paul SS, Dutta D, Ghosh A, Chattopadhyay K (2016) The non-native helical intermediate state may accumulate at low ph in the folding and aggregation landscape of the intestinal fatty acid binding protein. Biochemistry 55(32):4457–4468

Schuler B, Hofmann H (2013) Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. Curr Opin Struct Biol 23(1):36–47

Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK (2007) DisProt: the database of disordered proteins. Nucleic Acids Res 35(Database issue):D786–D793

Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, Seidel CA (2010) Accurate single-molecule FRET studies using multiparameter fluorescence detection. Methods Enzymol 475:455–514

Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) Alpha-synuclein in Lewy bodies. Nature 388(6645):839–840

Stohr J, Watts JC, Mensinger ZL, Oehler A, Grillo SK, DeArmond SJ, Prusiner SB, Giles K (2012) Purified and synthetic Alzheimer's amyloid beta (Abeta) prions. Proc Natl Acad Sci U S A 109 (27):11025–11030

Tashiro M, Kojima M, Kihara H, Kasai K, Kamiyoshihara T, Uéda K, Shimotakahara S (2008) Characterization of fibrillation process of α-synuclein at the initial stage. Biochem Biophys Res Commun 369(3):910–914

Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27(10):527–533

Trexler AJ, Rhoades E (2013) Function and dysfunction of alpha-synuclein: probing conformational changes and aggregation by single molecule fluorescence. Mol Neurobiol 47(2):622–631

Uversky VN (2011) Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol 43 (8):1090–1103

Veldhuis G, Segers-Nolten I, Ferlemann E, Subramaniam V (2009) Single-molecule FRET reveals structural heterogeneity of SDS-bound alpha-synuclein. Chembiochem 10(3):436–439

Widengren J, Kudryavtsev V, Antonik M, Berger S, Gerken M, Seidel CA (2006) Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. Anal Chem 78(6):2039–2050

Winkler J, Seybert A, Konig L, Pruggnaller S, Haselmann U, Sourjik V, Weiss M, Frangakis AS, Mogk A, Bukau B (2010) Quantitative and spatio-temporal features of protein aggregation in Escherichia coli and consequences on protein quality control and cellular ageing. EMBO J 29 (5):910–923

Ziv G, Haran G (2009) Protein folding, protein collapse, and tanford's transfer model: lessons from single-molecule FRET. J Am Chem Soc 131(8):2942–2947

Protein Misfolding and Neurodegenerative Diseases

11

Anand Narayan Singh, Nivedita Saxena, and Manish Kumar Verma

Abstract

A protein must fold into a three-dimensional structure and maintain its tertiary/quaternary structure for proper function. To control all the vital cellular activities, a protein must fold in a definite manner. Folding of protein is a quite complex phenomenon and is susceptible to errors resulting in a misfolded protein which can be lethal. There are a large number of neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), Huntington' disease (HD), Creutzfeldt–Jakob's disease, Cystic fibrosis, Gaucher's disease, prions disease, polyglutamine disease (PGD), and many other diseases caused due to the misfolded proteins. The mutation may cause protein misfolding and allow non-functional protein to accumulate and form amyloids. As a result, the pathological condition is fundamentally rooted. Molecular chaperones and ubiquitin protease system are the two main structures that perform crucial role in cellular function and survival. These two protein quality control systems specify and target the misfolded protein to a degradation pathway. If this system fails, an error in folding occurs leading to life-threatening diseases.

Keywords

Parkinson's disease · Alzheimer's disease · Cystic fibrosis · Gaucher's disease · Protein misfolding · Molecular chaperones · HSP70s · HSP40s

A. N. Singh (✉) · N. Saxena · M. K. Verma
Department of Biochemistry, GSVM Medical College, Kanpur, Uttar Pradesh, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_11

11.1 Introduction

Life relies upon the functioning of proteins, which depends on getting the correct folding of proteins. Proteins play many roles inside living cells including replication of genetic material, catalyzing metabolic reactions, cellular structure maintenance, cellular signaling, immune responses, cell adhesion, cell cycle, stimulus response, and in the transportation of molecules from one place to another. The unique conformation of a protein involves folding of its polypeptide chain in a characteristic three-dimensional structure.

Misfolding of proteins may occur due to various reasons such as improper posttranslational modifications, unfavorable environmental conditions, various kinds of mutations, oxidative stress, and error in trafficking (Uversky 2014). Aggregates of proteins are the consequences of misfolded polypeptides that distract from the regulatory points which are chaperones, heat shock proteins (HSP90, HSP70, etc.), and proteasome system. Protein aggregation is a slow phenomenon, early diagnosis of an individual that neurodegenerative disease will occur later in life is critical. Aggregations of misfolded proteins affect neuronal connectivity and plasticity and regulate cell death signaling pathways, which indicate many neurodegenerative diseases (Bence et al. 2001; Muchowski and Wacker 2005).

11.1.1 Basic Structure of Proteins

α-Carboxyl group (–COOH) of an amino acid joins with the α-amino group ($–NH_2$) of another amino acid to form a peptide bond (CO–NH) bridge. Proteins are formed by the polymerization of amino acids through peptide linkage; a dipeptide is formed by the combination of two amino acids; three and four amino acids form tripeptide and tetrapeptide, respectively, and over four amino acids combine and form an oligopeptide; a combination of 10–50 amino acids is called as a polypeptide.

11.1.2 Primary Structure Regulates Functional and Biological Activity of Protein

The specific primary structure of a protein will spontaneously form its natural three-dimensional conformations or structure. A higher level of an organization depends on the basic primary structure of the protein (Anfinsen's dogma). A three-dimensional structure of a functional protein also determines its biological function. This structural conformation provides and maintains the functional characteristics of a protein. In the three-dimensional structure of proteins, the hydrophilic polar charged residues are seen on the outer surface, and the non-polar hydrophobic residues remain inside, out of contact with water. A single amino acid change, i.e., mutation in the linear sequence, may have profound biological effects on the function, e.g., in normal hemoglobin (HbA); the sixth amino acid in the β-chain is glutamic acid which is replaced by valine in sickle cell anemia (HbS).

11.2 Protein Folding

For the maintenance of proteome integrity and health of a cell, all the processes should be in a proper way such as protein synthesis, folding and degradation, and the formation of a particular protein should be sharply regulated. Protein folding is a physical phenomenon by which a polypeptide chain folds into distinguished and functional three-dimensional conformation and ultimately forms a random coiled structure with a low-energy state.

The protein folding is a difficult and potentially perilous phenomenon (Luheshi et al. 2008). Folding of a protein in a cell involves sequenced and ordered pathways. With the continuation of the folding of a peptide, secondary structures are formed by the combination of hydrophobic groups together with the release of one molecule of water, i.e., exhibiting hydrophobic effect. Thus, the secondary structure is formed with other additional events, and then the tertiary structure is stabilized. In this manner, each substituent of secondary or super-secondary structure facilitates proper folding by directing the folding phenomenon toward the native structure of proteins (Fig. 11.1). Proteins fold into their active native state when they are released from the ribosome and also when they repeatedly unfold and refold during their lifetime (Bai et al. 1995; Bai and Englander 1996).

11.3 Protein Quality Control System (Assisted Folding)

The protein quality control (PQC) system organizes the proper folding of proteins. This PQC system eliminates misfolded and damaged proteins and prevents aggregation of proteins before they may exert toxic effects. The major constituents of a

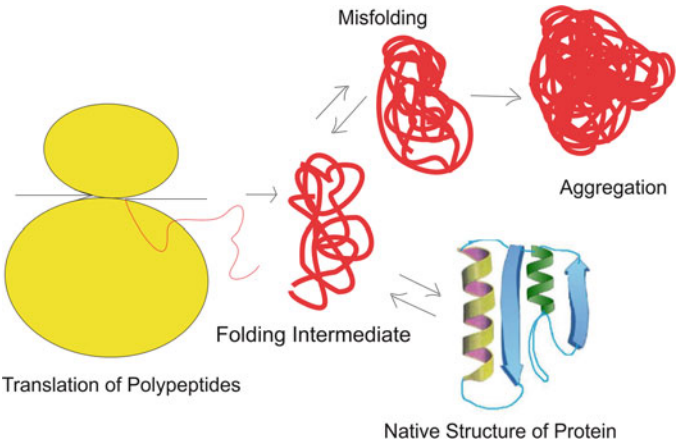


Fig. 11.1 Protein folding and misfolding. After translational process the polypeptide chain is released from ribosome moiety, folding of the polypeptide chain occurs and it can either form proper native conformation of protein or get misfolded and form aggregates

quality control system comprise molecular chaperones, ubiquitin-proteasome system (UPS), heat shock proteins (HSPs), and the unfolded protein response (UPR). These constituents of protein homeostasis or PQC system recognize misfolded or damaged proteins. Cells use this control system with auxiliary proteins to facilitate folding and direct it toward a productive and accurate conclusion. The proteins of the PQC system are as follows.

11.3.1 Chaperones

The chaperones represent a highly complex and modular molecular system. Chaperones act by protecting the hydrophobic residues of amino acids that are exposed in their non-native state but are hidden in their native conformation. Chaperones collaborate with the organized phenomenon of protein degradation and help them degrade in a distinct manner. Different chaperones escort the nascent polypeptide chain on the ribosomes at the translation process and their proper folding to escape protein from misfolding and aggregation. Some particular proteins of chaperones also function in the endoplasmic reticulum (ER) such as BiP, calnexin, calreticulin, and ERp 57, and identify misfolded or abnormal proteins and assist them holding in the ER, permitting particularly properly folded proteins to get the cytoplasm (Swanton et al. 2003). Chaperones are often activated by toxic chemicals, heat shock, oxidative stress, or inflammation (Garrido et al. 2001). During aging, the reason for the decrease in the protein folding process might be the improper balance between chaperone proteins and activity in neurons (Castro et al. 2018).

11.3.2 Heat Shock Proteins

Heat shock proteins (HSPs) are such types of proteins that are created by cells against stressful conditions. Many HSP proteins function as chaperone by stabilizing nascent proteins to ensure proper folding or by assisting them to refold that are destroyed using cellular stress. Heat shock proteins are categorized according to their molecular weight. According to their size, HSP40, HSP60, HSP70, and HSP90 are the categories of heat shock proteins on the order of 40, 60, 70, and 90 kDa, respectively.

In eukaryotes, HSP90 being a requisite protein is a crucial controller of the protein folding phenomenon in the cell (Jackson 2013). In a cell, Hsp90 combines with various substitute proteins, known as clients. HSP90 facilitates proper maturation, activation, and degradation of these HSP clients. Almost all the HSP90 clients, such as tau, kinases, synuclein, huntingtin, transcription factors, steroid hormone receptors, and E3 ubiquitin ligases, are unsimilar in their structure (Picard 2002; Shelton et al. 2017; Daturpalli et al. 2013; He et al. 2017; Hahn 2009). These proteins control different cellular processes including folding, degradation, cellular differentiation, cell growth, chromatin remodeling, and trafficking (Jackson 2013). The structure of the HSP70 system is shown in Fig. 11.2.

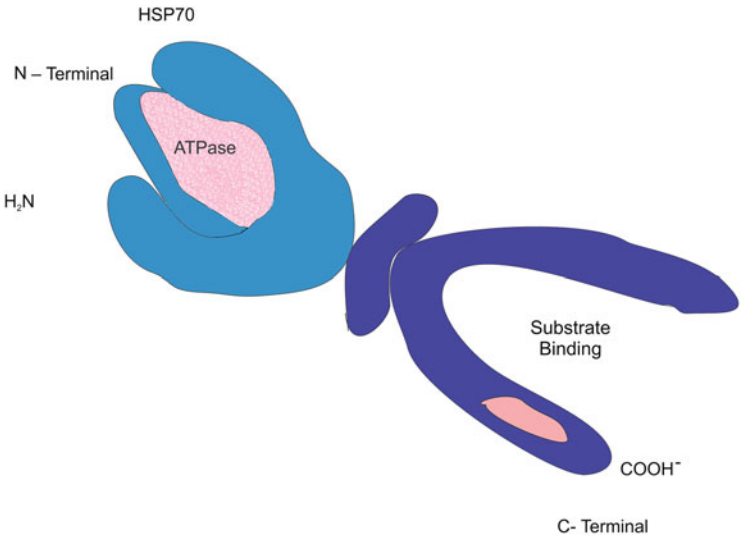


Fig. 11.2 HSP70 system, the structure of HSP 70: it contains an N terminal ATPase domain that binds with ATP and hydrolyses ATP to ADP and other sites of substrate-binding domain, contains β-sheet sub-domain which encloses the peptide backbone of the substrate. C-terminal domain, which is an α-helical structure, functions as a lid for the substrate-binding domain. When an HSP 70 protein binds with ATP, the lid gets open and peptides are bound and released rapidly. When an HSP 70 protein binds with ADP, the lid gets closed and peptides bind tightly with the substrate-binding domain.

11.3.2.1 Role as Chaperones

Various heat shock proteins function as intracellular chaperones. These HSP proteins are engaged in activities like protein–protein interactions such as folding and helping in the proper protein structure formation and eradicate the unwanted protein aggregation. Within the cell, HSPs also take part in transporting proteins across membranes. HSP40 transfers a nascent amino acid chain (unfolded) to HSP70, which takes the amino acid and facilitate its folding into an organized functional structure and then releases it. HSP60 grabs a nascent amino acid chain that has vanished its original structure and holds it. Chemical interactions in the protein structural conformation give the protein its accurately folded shape. HSP90 receives folded proteins from other chaperone clients and merges them into a large structure of a protein, e.g., a cellular receptor.

11.3.3 Proteasomes

Proteasomes are proteins that regulate the amount of particularly targeted proteins and degrade misfolded proteins. Proteins that are decided to degrade, tagged with a small protein called ubiquitin. Proteasome put an end to thousands of short-lived, damaged, misfolded, or otherwise obsolete proteins and perform an important role in protein quality control and other vital processes in the cell. It requires ATP for

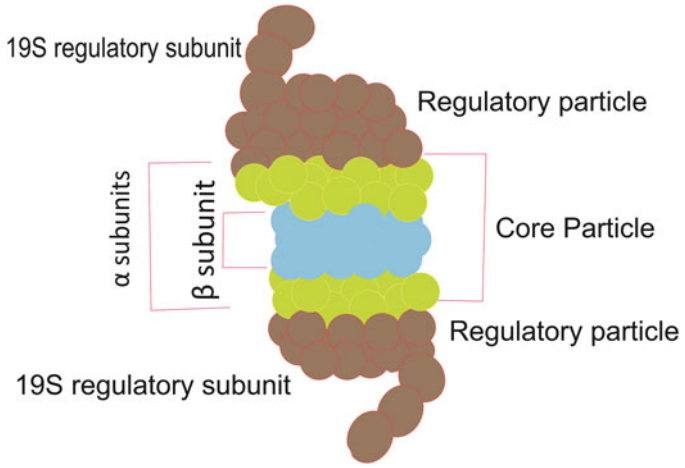


Fig. 11.3 Structure of a proteasome. A schematic diagram of a proteasome: it has a core particle (20S subunit) containing alpha and beta subunit and regulatory particles (19S regulatory subunit) consisting of the base and lid subcomplexes

metabolic energy needs. It is hollow and has openings at both sides for the entry of the protein to be processed and digested (Fig. 11.3). A human cell has about 30,000 proteasomes. This barrel-shaped structure can degrade practically all proteins to 7- to 9-amino-acid-long peptides.

The proteasome degrades undesirable cellular proteins by UPS, a multi-catalytic proteinase system. If a protein is not folded into its native conformation, then the chaperones trigger the misfolded proteins to be degraded or refolded with the help of the ubiquitin-proteasome pathway to avoid the formation of damaging or abnormal proteins. A misfolded polypeptide requires a multistep pathway for degradation and gets attach covalently with ubiquitin monomers. E1 or ubiquitin-activating enzyme activates the ubiquitin then transfers it to ubiquitin-conjugating enzyme, i.e., E2. Ultimately, enzyme ubiquitin ligase (E3) covalently combines ubiquitin with the protein. Thus, this sequence forms a polyubiquitin chain and targets the misfolded polypeptide chain to the proteasome for degradation (Glickman and Ciechanover 2002). This advance system of chaperones and units of the ubiquitin-proteasome system creates the most efficient cytosolic PQC system.

11.4 Mechanism and Intermediates in Protein Misfolding and Aggregation

Under stressful conditions such as high temperature, toxic chemicals or other factors, proteins mostly lost their identity in the cell. These factors may cause the misfolding of proteins and aggregates are formed (Ciechanover and Kwon 2015). These aggregates make an exertion on the cell since it causes major default regulation in

the metabolism of the cell (Cuanalo-Contreras et al. 2013). Protein misfolding, oligomerization, and aggregation are the main pathological abnormalities responsible to cause disease. Neurodegenerative diseases that debilitate our aging population, most prominently are AD, PD, amyotrophic lateral sclerosis, Huntington's disease, dementia with Lewy bodies, prion diseases, spinocerebellar ataxias, frontotemporal dementia, corticobasal degeneration, chronic traumatic encephalopathy and multiple system atrophy, type II diabetes, some types of cardiac diseases and cancer.

On the pathological basis, these diseases can be classified into two groups: loss of function and toxic gain of function. Improper protein functioning caused due to mutations may alter protein stability (forms a metastable protein) and cause degradation, as in cystic fibrosis and other metabolic defects comprise loss-of-function group (Sahni et al. 2015). The second group, i.e., toxic gain-of-function includes metastable proteins that aggregate and cause cellular toxicity.

11.4.1 Mechanism of Protein Misfolding

First, chaperones are expressed when a misfolding in protein is generated. This response of proteins or chaperones is named as the unfolded UPR in the endoplasmic reticulum (ER), whereas this response is known as the heat shock response (HSR) in the nucleus and cytosol. These responses are characterized as emergency responses to these stressful conditions; small alterations in protein homeostasis are addressed by these systems and play an important role in facilitating protein folding at the accurate place and help them to regain their accurate structure (Hartl et al. 2011). When it is confirmed that a misfolded protein is not to refold, then the PQC system, i.e., the proteasome, autophagy, and ER-associated degradation (ERAD) prohibited to correct and degrade these misfolded proteins (Nedelsky et al. 2008; Smith et al. 2011; Varshavsky 2012). Improper functioning in any of these pathways ultimately causes protein misfolding disease or neurodegenerative disease. Recent studies have shown that modulation of autophagy can be used for the treatment of amyloid diseases (Mputhia et al. 2019; Tripathi et al. 2019).

11.4.1.1 Aggregate Formation

In a cell, oligomers and aggregates are being formed when a maximum amount of misfolded protein is reached. These aggregated proteins ultimately form an amyloid-like structure, which eventually causes neurodegenerative disorders and cell death (Berke and Paulson 2003). The chief function of molecular chaperones is to assist folding by protecting the protein from misfolding, which may lead to aggregation. An α-helical structure is degraded and the simultaneous stabilization of β-sheets leads to the formation of an aggregate. Hydrophobic amino acid residues remain exposed in misfolded proteins and aggregate is formed having β-sheets. Most aggregates are amorphous in nature.

11.4.2 HSP70s and HSP40s in Protein Misfolding Diseases

HSP70s forms the largest fraction (27%) in the protein quality control system. The next frequently identified family is HSP40s (23%). HSP70s perform a basic quality control system and maintain functions of cells such as accurate folding of nascent polypeptides to avoid misfolding and aggregation. HSP70s prevent abnormal protein–protein interactions through the various cycles of ATP-binding and hydrolysis.

ATP-bounded HSP70 protein interacts with an unfolded amino acid chain. A more stable interaction is formed between the ADP-bound form of HSP70 and its substrate protein with the conversion of ATP to ADP. Then HSP70 releases this substrate protein (Mayer and Bukau 2005; Mayer 2013). In this manner, HSP70s play an important role in proteome maintenance and integrity for the stabilization of the native structure of a particular protein (Fig. 11.4).

HSP40, as co-chaperone, also performs an important role in the HSP70-HSP40 complex by activating ATP hydrolysis of HSP70 (Kampinga and Craig 2010; Kakkar et al. 2014) (Fig. 11.4). Any alteration in the function of HSP40 could change the folding capacity and HSP70 client specificity results in risk for the

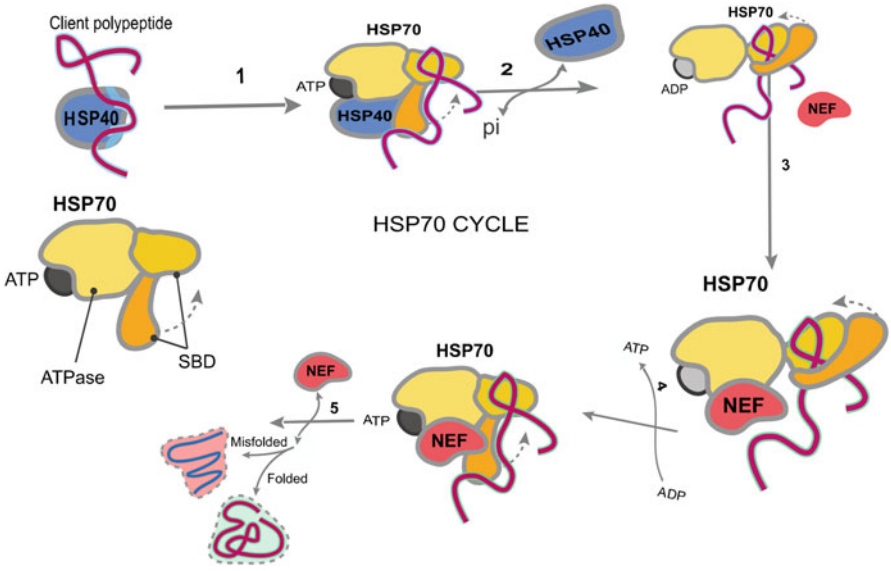


Fig. 11.4 HSP70 cycle. HSP40 ties to an amino acid chain or polypeptide chain (1) and joined with HSP70. (2) The HSP70 that is bounded with ATP communicates with an unfolded polypeptide chain through its substrate-binding domain (SBD). A progressively steady connection is formed between ADP-bound type of HSP70 and polypeptide after the hydrolysis of ATP to ADP invigorated with HSP40. (3) HSP70 interfaces with a nucleotide trade factor (NEF) and complex of the polypeptide is formed (4) It permits the conversion of ATP to ADP. (5) Detachment of both polypeptide and NEF from HSP70 takes place. On the other hand, if the polypeptide is not appropriately folded, the HSP70 cycle gets repeated

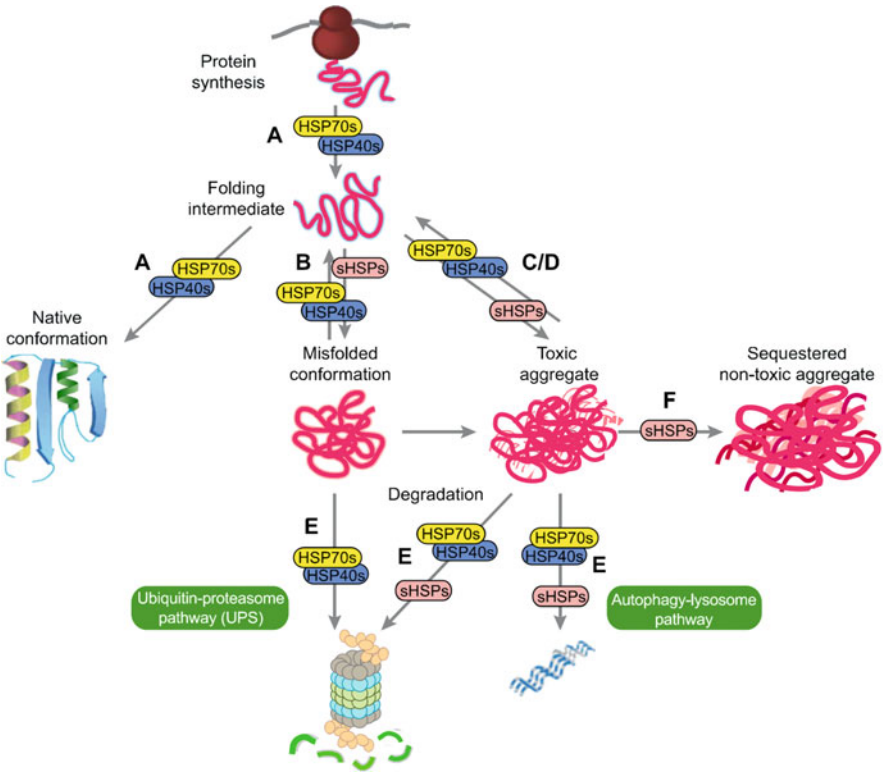


Fig. 11.5 Modification activities of HSP proteins in native and misfolding polypeptide. Both HSP70s and HSP40s together perform quality control process and assist folding process. (**a**) Misfolded proteins are refolded; (**b**) misfolding is arrested; (**c**) advancing proper folding; (**d**) misfolded polypeptides are degraded; (**e**) degradation of misfolded proteins through UPS and autophagy-lysosome pathway forming native conformation. (**f**) Toxic aggregates are sequestered and cells are protected by sHSPs

cellular proteostasis (Heldens et al. 2010) (Fig. 11.5). HSP40 proteins do not represent such a high state of conservation as in HSP70s. All HSP40s exhibit a 70-amino-acid motif, a J domain, that combines with HSP70s and activate ATPase enzyme (Kampinga and Craig 2010).

From 19 chaperone members, 32% of members are of the sHSP family which is also ATP-dependent. sHSPs are large and dynamic oligomers that attach with partially unfolded proteins. These sHSPs facilitate the binding and release of substrate protein without utilizing ATP (Bakthisaran et al. 2015). sHSPs forms a barrier to prevent abnormal protein interactions. In humans, about 10sHSP family members are known.

11.4.3 Energy Landscape in Protein Folding and Misfolding

Proteins consist of a specific and vast arrangement of folding at the interior sites that combine and form a thermodynamically stable three-dimensional structure, and for most of the proteins, free-energy gain of approximately -3 to -7 kcal mol^{-1} is needed for the proper folding of a protein compared to misfolded forms (Lindquist and Kelly 2011; Tripathi 2013). A certain order of reactions occurs between residues for the first step, i.e., nucleation condensation process that accelerates the folding process through various transition states indicated by interatomic interactions which also remains in the native structure of the protein (Fersht and Daggett 2002; Mayor et al. 2003). The free energy of a protein is determined based on its conformation and interactions present between the amino acid residues (Fig. 11.6). Small alterations in the polypeptide chain may deviate from the surface of the landscape, leading to the formation of new minimum free energy which ultimately results in a misfolded protein, prone to aggregation.

Local minimum energy is raised by the amino acid substitutions in its native state, thus forming the stabilization of structure more cumbersome, which increases the possibility of misfolding of protein and may cause aggregation. In the PQC system, molecular chaperones participate in the rearrangement of these structures by the formation of interactions with other proteins or by exposure of hydrophobic amino acid residues to the outer surface and a higher level of free energy and entropy are attained and a new folding path of a particular protein can start.

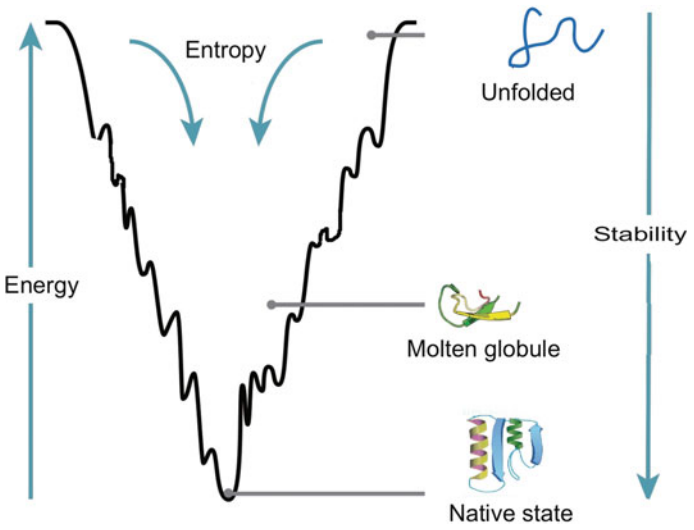


Fig. 11.6 Energy landscape of protein folding and misfolding. The energy landscape represents how proteins fold into their native structure by minimizing free energy

11.5 Protein Quality Control System in Particular Organelles and Associated Misfolding Diseases

Inside cells, protein folding occurs in the environment differs from that of a dilute buffer solution used in in vitro experiments. The phenomenon of protein folding takes place in various distinct organelles of a cell such as cytosol, mitochondria, ER, and peroxisomes. The cellular organelles (cytoplasm, ER, nucleus, etc.) containing proteins, membranes, and DNA; the level and heterogeneity of biomolecules may differ according to the compartment in which it is present. This large number of population of organelles results in increased viscosity, volume effects, and the amplified opportunity for specific and nonspecific intermolecular interactions. These organelles have a different chemical nature and may cause various protein folding issues that each cell must arrest and address. A list of protein misfolding diseases, related protein, and genes are given in Table 11.1.

11.5.1 Protein Quality Control System in the Cytosol

Most of the proteins undergo proper protein folding process in the cytosol when they release from the ribosome "quality control checkpoints" and then interact with chaperones and HSP. After releasing from the ribosome, these newly formed proteins are guarded by chaperones and other proteins like Hsp40, Hsp70, prefoldin, TCP-1 ring complex (TRiC), and nascent-polypeptide-associated complex (NAC) and form a competent folded protein until released from the ribosome. In the cytosol, various small proteins complete their folding process without any alliance, whereas a part of the cytosolic proteins complete their folding with the help of chaperones, e.g., Hsp90 and TRiC (Hartl and Hayer-Hartl 2002).

Among all the chaperones, TCP-1 ring complex (TRiC) is the most complicated double-ring chaperone of the cytosol. Each ring is formed from eight different subunits and made a large space in which the native polypeptide is formed. As the phenomenon of folding is completed, polypeptides are released into the cytosol (Spiess et al. 2004). If the chaperones are not working properly, misfolded proteins may be generated and aggregates are formed. In the cytosol, these protein aggregates may deposit beside the nuclear site and form an aggresome. These aggresomes may change into long amyloid fibrils. These are formed from the reverse transport of aggregated protein with microtubular tracks in a well-sequenced system of transportation (Kopito 2000). Amyloid fibrils are tube-like, long structures formed from the globular β-sheet structures (Dobson 2003). Deposition of fibrils may not be the initial toxic agent; however, the formation of soluble oligomers may be pathogenic (Bucciantini et al. 2002; Kayed et al. 2003; Cecchi et al. 2005; Mukai et al. 2005).

Table 11.1 Neurodegenerative diseases related to protein misfolding and their respective target proteins

S. no	Neurodegenerative disease	Target Protein/ Enzyme	Respective Gene	Reference
1.	Parkinson's disease	α-Synuclein	α-Synuclein, protein7, Pink1, cytosolic parkin, LRRK2, and VPS35	Carolyn (2011a, b, c); Hernandez et al. (2016)
2.	Amyotrophic lateral sclerosis (ALS)/Lou Gehrig's disease	Cu, Zn superoxide dismutase SOD1	SOD1	Siddique and Siddique 2001; Bruijn et al. (2004)
3.	Phenylketonuria	Phenyl alanine Cu, Zn, hydroxylase (PAH)	PAH	Carolyn (2011a, b, c)
4.	Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator, CFTR	Δ F508	De Boeck (2020); Kopito (1999)
5.	Emphysema	A-1-antitrypsin	SERPINA1	Perlmutter (2003)
6.	Medium-chain acyl CoA dehydrogenase deficiency (MCAD)	MCAD protein	K304E	Smon et al. (2018); Gregersen et al. (2004)
7.	Alzheimer's disease	Amyloid β-protein (Aβ) and microtubule-associated protein (tau), Presenilin protein	APP, PSEN1, or PSEN2	Carolyn (2011a, b, c)
8.	Huntington's disease	Huntingtin	HTT into mHTT	Carolyn (2011a, b, c); Bates (2003)
9.	Creutzfeldt–Jakob disease (CJD)	Prion protein	PrPc into PrPsc	NINDS (2020)
10.	Sickle cell anemia	HbS	Hemoglobin-Beta	Suzanne (2008)
11.	Nephrogenic diabetes insipidus (NDI)	Aquaporin-2/ vasopressin (AVPR2 gene)	AVPR2 or AQP2	Knoers and Lemmink 2000; Wildin et al. (1994)
12.	Retinitis pigmentosa (inherited blindness)	Rhodopsin	Thr58Arg, Pro347Leu, Pro347Ser, Ile-255	Huang et al. (2017); Dryja et al. (1990a, b); Berson et al. (1991a, b); Inglehearn et al. (1991)
13.	Fabry's disease (a lysosomal storage disorder)	α-Galactosidase	a-Gal A	Desnick et al. (2001); Eng et al. (1993); Ishii et al. (2002)

11.5.1.1 Cytosol-Associated Protein Misfolding Diseases

Parkinson's Disease

Parkinson's disease (PD) is the most known neurodegenerative disorder after AD and is indicated by slowed movement (bradykinesia), postural instability, muscular rigidity, and resting tremors. PD is slow but progressive. It involves degradation of neurotransmitter dopaminergic neurons in the substantia nigra region and also there is an accumulation of intracytoplasmic inclusion bodies in brain cells called Lewy bodies.

In PD, aggregation of proteins takes place in the brain, which ultimately leads to neurodegeneration (Moore et al. 2005). Most of PD are sporadic, but some may be due to inherited reasons such as a mutation in genes encoding synuclein, parkin, protein deglycase/PD protein7, Pink1, LRRK2, and VPS35 (Hernandez et al. 2016). These point mutations follow toxic gain-of-function due to the formation of α-synuclein cytosolic aggregates. These aggregates may be wild type or variant type.

These genes take part in the ubiquitination process. Alterations in α-synuclein and the pathogenesis of disease include deprivation in the function of PQC, which leads to α-synuclein aggregation with the generation of general oxidative stress which results in mitochondrial dysfunction (Bossy-Wetzel et al. 2004). In Lewy bodies, the Hsp90 level increases with α-synuclein. Hsp90 level increases in the brain of Parkinson's patients and correlates with an increased level of insoluble α-synuclein (Uryu et al. 2006). In vitro, recombinant Hsp90 suppresses α-synuclein to form aggregates without the use of ATP (Daturpalli et al. 2013). In PD, mitochondrial Pink1 and cytosolic parkin are also mutated and lead to inherited autosomal parkinsonism. Both Pink1 and parkin genes respond to mitochondrial abnormal proteins and safeguard the cell from the collection of destructed mitochondria (Scarffe et al. 2014).

Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic lateral sclerosis (ALS) is a disease that ultimately leads to neuronal death which controls voluntary muscles also known as motor neuron disease (MND) or Lou Gehrig's disease. This disease is distinguished by stiff muscles, twitching in muscles, and gradually weakness in limbs and slurred speech due to decreased muscle size. This disease is caused by alterations in Cu, Zn superoxide dismutase (SOD1) gene. Inhibition in the mutation of the superoxide radicals to hydrogen peroxide and oxygen is carried out by the SOD1 gene which guards the cell from oxidative damage (Bruijn et al. 2004). In the cytoplasm, aggregations of wild-type (normal) SOD1 protein are common in sporadic ALS (Brown and Al-Chalabi 2017). This mutation is a gain-of-function caused by the aggregation of the misfolded variant SOD1 protein. Mutant SOD1 gene can cause misfolding and aggregation of wild-type SOD1 in the neighboring neurons in a prion-like manner (Hardiman et al. 2017).

Phenylketonuria

Phenylketonuria (PKU) is known as an inborn error of phenylalanine metabolism (Al Hafid and Christodoulou 2015). PKU may be characterized by intellectual disability, seizures, behavioral changes, and mental disorders. PKU is caused by the defect in the phenylalanine hydroxylase (PAH) protein, which involves inaccurate folding of amino acid substitutions in PAH protein. Misfolding of a protein mainly occurs due to the amino acid substitutions in the PAH protein away from the enzyme's active site which renders the formation of active tetramer and ultimately cause degradation.

A baby from a mother having PKU may have many disturbances like heart problems, a small head, and low birth weight. PKU is inherited from the parents of an individual. Mutation in the PAH gene causes low enzyme phenylalanine hydroxylase levels. This leads to the conversion of dietary phenylalanine to a potentially harmful level. It is an autosomal recessive disorder. Two main types of PKU, i.e., classic PKU and variant PKU, may occur, based on which enzyme function remains.

11.5.2 Protein Quality Control System in the ER

ER is the major component of the secretory pathway. Ribosomal protein synthesis, co, and posttranslational modifications and protein folding are the main functions of the ER. A particular group of chaperones and protein folding enzymes are present in the ER lumen that mediates protein folding along with posttranslational modifications. Regarding this, the ER performs an essential role in the PQC system from the regulation of transfer of proteins from ER to the Golgi complex, particularly, for those proteins which have acquired their native conformation in the ER are released outside in a well-organized manner (Lee et al. 2004).

In a cell, calcium is stored in specialized compartments in the sarcoplasmic reticulum, golgi apparatus, lysosomes, and endosomes (Pinton et al. 2008; Rizzuto and Pozzan 2006). Calcium also gets reserved in the endoplasmic reticulum. For the protein folding and functioning of chaperones, an abundant amount of ER intraluminal calcium is needed. The ER also needs a high amount of energy for the protein folding process and a reduction in energy stops accurate folding of the protein. ATP in the form of energy is needed for accurate functioning of chaperones, for maintaining storage of Ca^{2+}, ER-associated degradation (ERAD) and for redox homeostasis (Malhotra and Kaufman 2010).

In the ER, there are two types of PQC system. All proteins are regulated at the primary PQC level. The primary PQC system is based on various factors such as identification of hydrophobic sequences, unpaired cysteine residues, immature glycans, and capacity of formation of protein aggregates. Primary PQC is composed of protein components, i.e., BiP, calnexin (CNX) and calreticulin (CRT), glucose-regulated protein 94 (GRP94), thiol-disulfide oxidoreductases, protein disulfide isomerase (PDI), and ERp57. Interactions of newly formed proteins with these components facilitate proper protein folding. Proteins which get misfolded initially

show a prolonged interaction with the PQC system but can aggregate without a systematic ER-associated degradation (Schröder and Kaufman 2005). The PQC system sorts out the protein for ERAD and transports it to the cytosol. In the cytosol, the UPS degrades these proteins (Tsai et al. 2002; Meusser et al. 2005). The secondary PQC system is regulated by cell-specific factors. This system may facilitate the secretion of a particular protein or various proteins. In mammalian cells, ER permits the secretion of various pathogenic transthyretin variants with wild-type efficiency. ERAD regulates the most highly destabilized variants and then to the specific tissues act to it (Sekijima et al. 2005).

11.5.2.1 ER-Associated Misfolding Diseases

Cystic Fibrosis

Cystic fibrosis is a disease caused by the thick mucus secretions in the lung and intestines (Muchowski 2002). Improper degradation of proteins may be the reason for the progression of more severe disease. This disease is caused due to the mutation in cystic fibrosis transmembrane conductance regulator (CFTR) which is a plasma membrane chloride channel. Within the cell membrane, CFTR protein carries two nucleotide-binding domains, 12 potential transmembrane domains, and a highly charged hydrophilic region, functions as a regulatory domain and can be observed by the analysis of amino acid sequence (Welch 2004).

Although various CFTR sequence mutations have been identified; within 1480 amino acids long polypeptide chain, one particular mutation is observed in 705 patients. The deletion of three nucleotides that are coding for a phenylalanine residue at position 508 (ΔF508 CFTR) takes place in this type of mutation (Kopito 1999). In the ER, the ΔF508 allele of the CFTR gene has been found as a trafficking mutation that prevents protein maturation and moving it for premature proteolysis (Kerem et al. 1989).

Emphysema

Pulmonary emphysema is a disease that comprises chronic obstructive pulmonary disease (COPD). This disease may be characterized by the gradual destruction of lung tissue, particularly thinning and damage of the alveoli or air sacs usually caused by significant exposure of toxic particles or gases (Rustagi et al. 2019; Fernandez-Bussy et al. 2018; Dunlap et al. 2019). Deficiency of α-1-antitrypsin causes hereditary emphysema which comprises ER-associated misfolding and rapid degradation of proteins. α-1 antitrypsin is secreted by hepatocytes. It is a plasma serine protease inhibitor that regulates the proteolytic activity of many enzymes. Under certain conditions, aggregation of this variant protein causes the damage of cells in the liver (Perlmutter 2003).

11.5.3 Protein Quality Control System in Mitochondria

In humans, mitochondria is an organelle of the cell where approximately 1500 proteins' folding and degradation processes occur (Taylor et al. 2003). The mitochondrial DNA encodes only 13 proteins of mitochondria. Mitochondria take these proteins inside it mainly in an unfolded form through pores present in the outer and inner membrane of mitochondria (Wiedemann et al. 2004).

Hsp70 and Hsp90 are the molecular chaperones that keep nascent mitochondrial proteins in an unfolded, import-competent state in the cytosol (Young et al. 2003). A mitochondrial Hsp70 client binds to the nascent polypeptide chain and is engaged in the translocation (Wiedemann et al. 2004). The mitochondrial PQC system of yeast and bacterial cells' molecular chaperones consists of mitochondrial Hsp70, the Hsp60/Hsp10 system, and a set of proteases with AAA+ domains that resemble with the proteasome and are present in the mitochondrial lumen or the inner membrane of mitochondria (Käser and Langer 2000).

11.5.3.1 Mitochondria-Associated Protein Misfolding Diseases

Medium Acyl CoA Dehydrogenase Deficiency

Medium-chain acyl-CoA dehydrogenase deficiency (MCAD) is a disorder of fatty acid oxidation in which an individual cannot break down medium-chain fatty acids into acetyl-CoA. Hypoglycemia and sudden death, most often caused due to fasting or vomiting, are the characteristic feature of MCAD. The MCAD enzyme takes part in the β-oxidation of fatty acid in mitochondria. An amino acid substitution, K304E, is found to conduct the disease in approximately 90% of the persons (Gregersen et al. 2004). MCAD protein gets folded through successive interactions with mitochondrial system HSP70 and the mitochondrial chaperone system HSP60.

Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disease that affects memory and causes impaired characteristics in language, visuospatial, and motor dysfunctions (Melis et al. 2015). Extracellular depositions of Aβ and neurofibrillary tangles (NFTs) are formed in the brain. In AD, cleavage of a 42-amino-acid β-amyloid peptide occurs (Games et al. 1995; Oltersdorf et al. 1989). An amyloid precursor, a membrane protein, which after cleavage by β-secretase produces a β-amyloid precursor peptide fragment. Another protease β-secretase further breaks this fragment and forms Aβ42 instead of Aβ40, which is amyloidogenic. Many abnormal clumps, i.e., amyloid plaques and tau-based NFTs, are formed in the present disease. Under normal conditions, when produced in lesser quantity, degradation of Aβ42 is the normal process of this peptide fragment. However, in some specific conditions, it forms extracellular aggregates and then amyloids are formed. A mouse model of this disease proves that the insoluble tau in neuronal cells changes the solubility of many proteins, causing damage to cellular homeostasis (Pace et al. 2018). Recently, it has been shown that interaction between the β-amyloid core and

tau facilitates cross-seeding (Tripathi and Khan 2020), and their synergistic effect disrupts neural circuits (Tripathi and Kalita 2019).

11.5.4 Other Associated Neurodegenerative Diseases

11.5.4.1 Huntington's Disease

Huntington's disease is an inherited disorder that implies progressive degeneration of nerve cells in the brain. Huntington's disease results in an individual functional inability in movement, thinking (cognitive), and cause psychiatric disorders. Most of Huntington's disease patients develop signs and symptoms in their age of 30s or 40s, but the disease may emerge earlier or later in life.

It is an autosomal dominant disorder. In Huntington's disease, CAG trinucleotide repeat expansion in exon 1 of the huntingtin gene (HTT) occurs. These expanded CAG trinucleotide repeats are translated into a series of continuous glutamine residues resulting in the formation of a polyglutamine tract or polyQ tract. This extended PolyQ HTT can aggregate and create inclusion bodies and ultimately cause neuronal degeneration of neurons (Gusella and MacDonald 2006). This altered form, i.e., mutant huntingtin (mHTT), increases the deterioration rate of specific types of neurons. Physical interaction is found between wild-type HTT or mutated HTT (mHTT) and HSP90. When HSP90 is inhibited, the interaction between these proteins is disarranged, and the HTT gene is broken through the ubiquitin-proteasome system (Baldo et al. 2012).

11.5.4.2 Creutzfeldt–Jacob Disease

Creutzfeldt–Jacob disease is an inherited autosomal dominant disease. It is the human Prion Disease responsible for 85% cases in which 10–15% of cases are familial. Creutzfeldt–Jakob disease (vCJD) is a structural conversion of wild-type protein PrPc into β-sheet-dominant PrPsc, resulting in misfolding and then aggregation (Prusiner 2001; Cohen et al. 1994; Goldfarb et al. 1992). The human prion protein, PrPc, is a small cellular surface glycoprotein where "c" denotes "cellular." PrPc becomes PrPsc, where "sc" implies "scrapie" after transformation, which is very infectious (Wille and Requena 2018). The physiological role of PrPc is still under inspection. It may be engaged in shielding against stress, copper homeostasis, and neuronal excitability. PrPc takes part in the regulation of myelin maintenance, cellular differentiation, proliferation, adhesion, and cell morphology. Control of the circadian rhythm, glucose homeostasis, immune function, and cellular iron uptake may involve PrPc (Castle and Gill 2017).

11.5.4.3 Sickle Cell Anemia

Sickle cell anemia is a first protein misfolding disease. It is also the first inherited disease with a known molecular mechanism. It is a disorder in which the amino acid valine is mutated by glutamine at the sixth position of the β-globin chain. It is a single nucleotide mutation (GAG codon changing to GTG) of the β-globin gene,

which ultimately results in that glutamic acid (E/Glu) being substituted by valine (V/Val) at position 6 (E6V substitution) (Suzanne 2008).

This variation causes intermolecular binding among adjacent molecules of hemoglobin, and stable long polymer fiber-like structures are formed (Galkin and Vekilov 2004). This mutation leads to the formation of mutant hemoglobin S (HbS), a deoxygenated state which is a stable fiber-like structure. This mutation ultimately converts the shape and rigidity of red blood cells and causes abnormality. In this disease, many β-pleated sheets accumulate and form amyloid plaques.

11.5.4.4 Nephrogenic Diabetic Insipidus

Nephrogenic diabetes insipidus (NDI) is also a misfolding disorder of an antidiuretic hormone vasopressin, which is a hormonal protein. In this disease, kidneys are unable to remove water from the urine. In this disease, kidneys are resisting the action of arginine vasopressin (Inaba et al. 2001). Hereditary nephrogenic diabetes insipidus can result from the mutation in at least two genes. Mostly, hereditary nephrogenic diabetes insipidus can be from mutations in the AVPR2 gene. Arginine vasopressin is encoded by the mutated AVPR2 gene in NDI (Wildin et al. 1994). Most of the remaining cases are caused by the mutations in the AQP2 gene. Both the genes i.e., AVPR2 and AQP2 provide instructions to make proteins that help to decide how much water is excreted in the urine.

11.5.4.5 Retinitis Pigmentosa

Retinitis pigmentosa (RP) is distinguished by night blindness in which loss of peripheral vision is accompanied by a loss in the central vision. It is the most common cause of inherited blindness with over 25 genetic loci identified. Mutations in the gene encoding rhodopsin cause this disease (Dryja et al. 1990a, b). Misfolded rhodopsin follows a gain of function that ultimately causes cell death. Misfolding in rhodopsin in the intradiscal, transmembrane, and cytoplasmic domains inhibits the translocation to the plasma membrane, and it stores in the ER and Golgi complex to cause the disease (Chapple et al. 2001).

One of the chief biochemical reason for RP in rhodopsin mutations is protein misfolding and the disarrangement of molecular chaperones (Senin et al. 2006). Most of the codon mutations associated with retinitis pigmentosa include Thr58Arg, Pro347Leu, Pro347Ser, and deletion of Ile255 (Berson et al. 1991a, b; Dryja et al. 1990a, b; Inglehearn et al. 1991).

11.5.4.6 Fabry's Disease

Fabry's disease or Anderson–Fabry's disease is a rare genetic disease. This disease can harm many parts of the body, including the kidneys, heart, and skin (Timothy and Elston 2011). It is characterized by a painful crisis, angiokeratomas, corneal dystrophy, and hypohydrosis (Perrot et al. 2002).

This disease is a lysosomal storage disease. It is an insufficiency in the activity of galactosidase A enzyme in lysosomes which results in the storage of glycosphingolipid globotriaosylceramide (Gb3). In Fabry's disease, missense mutations occur in the a-Gal A gene (GLA), but alternative splicing mutations and

deletions have also been found (Eng et al. 1993; Ishii et al. 2002). These mutant enzymes are misfolded and get recognized by the protein quality control system of ER and are degraded before reaching to lysosomes for sorting. Fabry's disease is caused particularly by those missense mutations that result in misfolding of a-Gal gene.

11.6 Conclusion

Protein misfolding-related neurodegenerative diseases also known as "conformational diseases" ultimately result from the misfolding of proteins. Conversion of an α-helix to intermolecular β-sheets ultimately forms aggregated conformations. This conformation is maintained by intermolecular interactions, leading to the generation of oligomers, proto-fibrils, and fibrils which then accumulate as amyloid structures in the affected cells. This accumulation can be intra- or extracellular in the CNS or the periphery. Alzheimer disease, Parkinson's disease, amyotrophic lateral sclerosis, Huntington's disease, dementia with Lewy bodies, prion diseases, spinocerebellar ataxias, frontotemporal dementia, corticobasal degeneration, chronic traumatic encephalopathy and multiple system atrophy, type II diabetes, and certain forms of heart disease and cancer are the diseases caused by misfolding of proteins, considered under the category of neurodegenerative diseases.

Molecular chaperones, HSPs, and proteasomes are the components of the PQC system which prevent aggregation and eliminate misfolded and damaged proteins before they may exert toxic effects. If this PQC system fails, it causes neurodegenerative diseases. These diseases are such debilitating diseases that are life threatening to humans. So, more work is needed to terminate these disorders.

References

Al Hafid N, Christodoulou J (2015) Phenylketonuria: a review of current and future treatments. Transl Pediatr 4(4):304–317

Bai Y, Englander SW (1996) Future directions in folding: the multi-state nature of protein structure. Proteins 24(2):145–151

Bai Y, Englander JJ, Mayne L, Milne JS, Englander SW (1995) Thermodynamic parameters from hydrogen exchange measurements. Methods Enzymol 259:344–356

Bakthisaran R, Tangirala R, Rao CM (2015) Small heat shock proteins: role in cellular functions and pathology. Biochim Biophys Acta 1854(4):291–319

Baldo B, Weiss A, Parker CN, Bibel M, Paganetti P, Kaupmann K (2012) A screen for enhancers of clearance identifies huntingtin as a heat shock protein 90 (Hsp90) client protein. J Biol Chem 287(2):1406–1414

Bates G (2003) Huntingtin aggregation and toxicity in Huntington's disease. Lancet 361 (9369):1642–1644

Bence NF, Sampat RM, Kopito RR (2001) Impairment of the ubiquitin-proteasome system by protein aggregation. Science 292:1552–1555

Berke SJ, Paulson HL (2003) Protein aggregation and the ubiquitin proteasome pathway: gaining the UPPer hand on neurodegeneration. Curr Opin Genet Dev 13(3):253–261

Berson EL, Rosner B, Sandberg MA, Dryja TP (1991a) Ocular findings in patients with autosomal dominant retinitis pigmentosa and a rhodopsin gene defect (Pro-23-His). Arch Ophthalmol 109 (1):92–101
Berson EL, Rosner B, Sandberg MA, Weigel-DiFranco C, Dryja TP (1991b) Ocular findings in patients with autosomal dominant retinitis pigmentosa and rhodopsin, proline-347-leucine. Am J Ophthalmol 111(5):614–623
Bossy-Wetzel E, Schwarzenbacher R, Lipton SA (2004) Molecular pathways to neurodegeneration. Nat Med 10(Suppl):S2–S9
Brown RH, Al-Chalabi A (2017) Amyotrophic lateral sclerosis. N Engl J Med 377(2):162–172
Bruijn LI, Miller TM, Cleveland DW (2004) Unraveling the mechanisms involved in motor neuron degeneration in ALS. Annu Rev Neurosci 27:723–749
Bucciantini M, Giannoni E, Chiti F et al (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 416(6880):507–511
Carolyn L (2011a) Crankshaw, neurodegenerative diseases: Alzheimer's disease. Biofiles 7(2):4–8
Carolyn L (2011b) Crankshaw, neurodegenerative diseases: Huntington's disease. Biofiles 7 (2):9–14
Carolyn L (2011c) Crankshaw, neurodegenerative diseases: Parkinson's disease. Biofiles 7 (2):16–20
Castle AR, Gill AC (2017) Physiological functions of the cellular prion protein. Front Mol Biosci 4:19
Castro JP, Wardelmann K, Grune T, Kleinridders A (2018) Mitochondrial chaperones in the brain: safeguarding brain health and metabolism? Front Endocrinol (Lausanne) 9:196
Cecchi C, Baglioni S, Fiorillo C et al (2005) Insights into the molecular basis of the differing susceptibility of varying cell types to the toxicity of amyloid aggregates. J Cell Sci 118 (Pt 15):3459–3470
Chapple JP, Grayson C, Hardcastle AJ, Saliba RS, van der Spuy J, Cheetham ME (2001) Unfolding retinal dystrophies: a role for molecular chaperones? Trends Mol Med 7(9):414–421
Ciechanover A, Kwon YT (2015) Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. Exp Mol Med 47(3):e147
Cohen FE, Pan KM, Huang Z, Baldwin M, Fletterick RJ, Prusiner SB (1994) Structural clues to prion replication. Science 264(5158):530–531
Cuanalo-Contreras K, Mukherjee A, Soto C (2013) Role of protein misfolding and proteostasis deficiency in protein misfolding diseases and aging. Int J Cell Biol 2013:638083
Daturpalli S, Waudby CA, Meehan S, Jackson SE (2013) Hsp90 inhibits α-synuclein aggregation by interacting with soluble oligomers. J Mol Biol 425(22):4614–4628
De Boeck K (2020) Cystic fibrosis in the year 2020: a disease with a new face. Acta Paediatr https://doi.org/10.1111/apa.15155
Desnick R, Ioannou Y, Eng C (2001) Alpha-Galactosidase A deficiency: Fabry disease. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) The metabolic and molecular bases of inherited disease, 8th edn. McGraw-Hill, New York, pp 3733–3774
Dobson CM (2003) Protein folding and misfolding. Nature 426(6968):884–890
Dryja TP, McGee TL, Hahn LB et al (1990a) Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. N Engl J Med 323(19):1302–1307
Dryja TP, McGee TL, Reichel E et al (1990b) A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. Nature 343(6256):364–366
Dunlap DG, Semaan R, Riley CM, Sciurba FC (2019) Bronchoscopic device intervention in chronic obstructive pulmonary disease. Curr Opin Pulm Med 25(2):201–210
Eng CM, Resnick-Silverman LA, Niehaus DJ, Astrin KH, Desnick RJ (1993) Nature and frequency of mutations in the alpha-galactosidase A gene that cause Fabry disease. Am J Hum Genet 53 (6):1186–1197
Fernandez-Bussy S, Labarca G, Herth FJF (2018) Bronchoscopic lung volume reduction in patients with severe emphysema. Semin Respir Crit Care Med 39(6):685–692

Fersht AR, Daggett V (2002) Protein folding and unfolding at atomic resolution. Cell 108 (4):573–582
Galkin O, Vekilov PG (2004) Mechanisms of homogeneous nucleation of polymers of sickle cell anemia hemoglobin in deoxy state. J Mol Biol 336(1):43–59
Games D, Adams D, Alessandrini R et al (1995) Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein. Nature 373(6514):523–527
Garrido C, Gurbuxani S, Ravagnan L, Kroemer G (2001) Heat shock proteins: endogenous modulators of apoptotic cell death. Biochem Biophys Res Commun 286(3):433–442
Glickman MH, Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiol Rev 82(2):373–428
Goldfarb LG, Brown P, Haltia M et al (1992) Creutzfeldt-Jakob disease cosegregates with the codon 178Asn PRNP mutation in families of European origin. Ann Neurol 31(3):274–281
Gregersen N, Bross P, Andresen BS (2004) Genetic defects in fatty acid beta-oxidation and acyl-CoA dehydrogenases. Molecular pathogenesis and genotype-phenotype relationships. Eur J Biochem 271(3):470–482
Gusella JF, MacDonald ME (2006) Huntington's disease: seeing the pathogenic process through a genetic lens. Trends Biochem Sci 31(9):533–540
Hahn JS (2009) The Hsp90 chaperone machinery: from structure to drug development. BMB Rep 42(10):623–630
Hardiman O, Al-Chalabi A, Chio A et al (2017) Amyotrophic lateral sclerosis. Nat Rev Dis Primers 3:17085
Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295(5561):1852–1858
Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475(7356):324–332
He WT, Xue W, Gao YG et al (2017) HSP90 recognizes the N-terminus of huntingtin involved in regulation of huntingtin aggregation by USP19. Sci Rep 7(1):14797
Heldens L, Dirks RP, Hensen SM et al (2010) Co-chaperones are limiting in a depleted chaperone network. Cell Mol Life Sci 67(23):4035–4048
Hernandez DG, Reed X, Singleton AB (2016) Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. J Neurochem 139:59–74
Huang L, Zhang Q, Huang X et al (2017) Mutation screening in genes known to be responsible for retinitis Pigmentosa in 98 small Han Chinese families. Sci Rep 7(1):1948
Inaba S, Hatakeyama H, Taniguchi N, Miyamori I (2001) The property of a novel v2 receptor mutant in a patient with nephrogenic diabetes insipidus. J Clin Endocrinol Metab 86(1):381–385
Inglehearn CF, Bashir R, Lester DH, Jay M, Bird AC, Bhattacharya SS (1991) A 3-bp deletion in the rhodopsin gene in a family with autosomal dominant retinitis pigmentosa. Am J Hum Genet 48(1):26–30
Ishii S, Nakao S, Minamikawa-Tachino R, Desnick RJ, Fan JQ (2002) Alternative splicing in the alpha-galactosidase A gene: increased exon inclusion results in the Fabry cardiac phenotype. Am J Hum Genet 70(4):994–1002
Jackson SE (2013) Hsp90: structure and function. Top Curr Chem 328:155–240
Kakkar V, Meister-Broekema M, Minoia M, Carra S, Kampinga HH (2014) Barcoding heat shock proteins to human diseases: looking beyond the heat shock response. Dis Model Mech 7 (4):421–434
Kampinga HH, Craig EA (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol 11(8):579–592
Käser M, Langer T (2000) Protein degradation in mitochondria. Semin Cell Dev Biol 11 (3):181–190
Kayed R, Head E, Thompson JL et al (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science 300(5618):486–489
Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245:1073–1080

Knoers N, Lemmink H. Hereditary Nephrogenic Diabetes Insipidus. 2000 Feb 12 [Updated 2020 Feb 27]
Kopito RR (1999) Biosynthesis and degradation of CFTR. Physiol Rev 79(1 Suppl):S167–S173
Kopito RR (2000) Aggresomes, inclusion bodies and protein aggregation. Trends Cell Biol 10 (12):524–530
Lee MC, Miller EA, Goldberg J, Orci L, Schekman R (2004) Bi-directional protein transport between the ER and Golgi. Annu Rev Cell Dev Biol 20:87–123
Lindquist SL, Kelly JW (2011) Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases: progress and prognosis. Cold Spring Harb Perspect Biol 3(12):a004507
Luheshi LM, Crowther DC, Dobson CM (2008) Protein misfolding and disease: from the test tube to the organism. Curr Opin Chem Biol 12(1):25–31
Malhotra JD, Kaufman RJ (2010) In: Ramirez-Alvarado M, Kelly JW, Dobson CM (eds) Endoplasmic reticulum stress and oxidative stress: mechanisms and link to disease. Wiley and Sons Publication, New Jersey, pp 47–72
Mayer MP (2013) Hsp70 chaperone dynamics and molecular mechanism. Trends Biochem Sci 38 (10):507–514
Mayer MP, Bukau B (2005) Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci 62(6):670–684
Mayor U, Guydosh NR, Johnson CM et al (2003) The complete folding pathway of a protein from nanoseconds to microseconds. Nature 421(6925):863–867
Melis V, Zabke C, Stamer K, Magbagbeolu M, Schwab K, Marschall P, Veh RW, Bachmann S, Deiana S, Moreau PH et al (2015) Different pathways of molecular pathophysiology underlie cognitive and motor tauopathy phenotypes in transgenic models for Alzheimer's disease and frontotemporal lobar degeneration. Cell Mol Life Sci 72:2199–2222
Meusser B, Hirsch C, Jarosch E, Sommer T (2005) ERAD: the long road to destruction. Nat Cell Biol 7(8):766–772
Moore DJ, West AB, Dawson VL, Dawson TM (2005) Molecular pathophysiology of Parkinson's disease. Annu Rev Neurosci 28:57–87
Mputhia Z, Hone E, Tripathi T, Sargeant T, Martins R, Bharadwaj P (2019) Autophagy modulation as a treatment of amyloid diseases. Molecules 24(18):3372
Muchowski PJ (2002) Protein misfolding, amyloid formation, and neurodegeneration: a critical role for molecular chaperones? Neuron 35(1):9–12
Muchowski PJ, Wacker JL (2005) Modulation of neurodegeneration by molecular chaperones. Nat Rev Neurosci 6:11–22
Mukai H, Isagawa T, Goyama E et al (2005) Formation of morphologically similar globular aggregates from diverse aggregation-prone proteins in mammalian cells. Proc Natl Acad Sci U S A 102(31):10887–10892
Nedelsky NB, Todd PK, Taylor JP (2008) Autophagy and the ubiquitin-proteasome system: collaborators in neuro protection. Biochim Biophys Acta 1782(12):691–699
NINDS (2020) Creutzfeldt–Jakob disease fact sheet. National Institute of Neurological Disorders and Stroke (NINDS). http://www.ninds.nih.gov/disorders/cjd/detail_cjd.htm#264203058. Accessed 2 Feb 2020
Oltersdorf T, Fritz LC, Schenk DB et al (1989) The secreted form of the Alzheimer's amyloid precursor protein with the Kunitz domain is protease nexin-II. Nature 341(6238):144–147
Pace MC, Xu G, Fromholt S et al (2018) Changes in proteome solubility indicate widespread proteostatic disruption in mouse models of neurodegenerative disease. Acta Neuropathol 136 (6):919–938
Perlmutter DH (2003) Alpha1-antitrypsin deficiency: liver disease associated with retention of a mutant secretory glycoprotein in the endoplasmic reticulum. Methods Mol Biol 232:39–56
Perrot A, Osterziel KJ, Beck M, Dietz R, Kampmann C (2002) Fabry disease: focus on cardiac manifestations and molecular mechanisms. Herz 27(7):699–702

Picard D (2002) Heat-shock protein 90, a chaperone for folding and regulation. Cell Mol Life Sci 59 (10):1640–1648
Pinton P, Giorgi C, Siviero R, Zecchini E, Rizzuto R (2008) Calcium and apoptosis: ER-mitochondria $Ca2^{+}$ transfer in the control of apoptosis. Oncogene 27:6407–6418
Prusiner SB (2001) Shattuck lecture—neurodegenerative diseases and prions. N Engl J Med 344 (20):1516–1526
Rizzuto R, Pozzan T (2006) Microdomains of intracellular Ca^{2+}: molecular determinants and functional consequences. Physiol Rev 86:369–408
Rustagi N, Singh S, Dutt N et al (2019) Efficacy and safety of stent, valves, vapour ablation, coils and sealant therapies in advanced emphysema: a meta-analysis. Turk Thorac J 20(1):43–60
Sahni N, Yi S, Taipale M et al (2015) Widespread macromolecular interaction perturbations in human genetic disorders. Cell 161(3):647–660
Scarffe LA, Stevens DA, Dawson VL, Dawson TM (2014) Parkin and PINK1: much more than mitophagy. Trends Neurosci 37(6):315–324
Schröder M, Kaufman RJ (2005) The mammalian unfolded protein response. Annu Rev Biochem 74:739–789
Sekijima Y, Wiseman RL, Matteson J et al (2005) The biological and chemical basis for tissue-selective amyloid disease. Cell 121(1):73–85
Senin II, Bosch L, Ramon E et al (2006) Ca2+/recoverin dependent regulation of phosphorylation of the rhodopsin mutant R135L associated with retinitis pigmentosa. Biochem Biophys Res Commun 349(1):345–352
Shelton LB, Koren J 3rd, Blair LJ (2017) Imbalances in the Hsp90 chaperone machinery: implications for Tauopathies. Front Neurosci 11:724
Siddique N, Siddique T. Amyotrophic Lateral Sclerosis Overview. 2001 Mar 23 [Updated 2019 Oct 3]
Smith MH, Ploegh HL, Weissman JS (2011) Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. Science 334(6059):1086–1090
Smon A, Groselj U, Debeljak M et al (2018) Medium-chain acyl-CoA dehydrogenase deficiency: two novel ACADM mutations identified in a retrospective screening. J Int Med Res 46 (4):1339–1348
Spiess C, Meyer AS, Reissmann S, Frydman J (2004) Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. Trends Cell Biol 14(11):598–604
Suzanne C (2008) Genetic mutation. Nat Edu 1(1):187
Swanton E, High S, Woodman P (2003) Role of calnexin in the glycan-independent quality control of proteolipid protein. EMBO J 22(12):2948–2958
Taylor SW, Fahy E, Ghosh SS (2003) Global organellar proteomics. Trends Biotechnol 21 (2):82–88
Timothy WJ, Elston BD (2011) Andrews' diseases of the skin, 11th edn. Saunders Elsevier, London
Tripathi T (2013) Calculation of thermodynamic parameters of protein unfolding using far-ultraviolet circular dichroism. J Proteins Proteomics 4(2):85–91
Tripathi T, Kalita P (2019) Synergistic effect of amyloid-β and tau disrupts neural circuits. ACS Chem Neurosci 10(3):1129–1130
Tripathi T, Khan H (2020) Direct interaction between the β-amyloid core and tau facilitates cross-seeding: a novel target for therapeutic intervention. Biochemistry 59(4):341–342
Tripathi T, Kalita P, Martins R, Bharadwaj P (2019) Autophagy promotes memory formation. ACS Chem Neurosci 10(8):3337–3339
Tsai B, Ye Y, Rapoport TA (2002) Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. Nat Rev Mol Cell Biol 3(4):246–255
Uryu K, Richter-Landsberg C, Welch W et al (2006) Convergence of heat shock protein 90 with ubiquitin in filamentous alpha-synuclein inclusions of alpha-synucleinopathies. Am J Pathol 168(3):947–961
Uversky VN (2014) The triple power of D^{3}: protein intrinsic disorder in degenerative diseases. Front Biosci (Landmark Ed) 19:181–258

Varshavsky A (2012) The ubiquitin system, an immense realm. Annu Rev Biochem 81:167–176
Welch WJ (2004) Role of quality control pathways in human diseases involving protein misfolding. Semin Cell Dev Biol 15(1):31–38
Wiedemann N, Frazier AE, Pfanner N (2004) The protein import machinery of mitochondria. J Biol Chem 279(15):14473–14476
Wildin RS, Antush MJ, Bennett RL, Schoof JM, Scott CR (1994) Heterogeneous AVPR2 gene mutations in congenital nephrogenic diabetes insipidus. Am J Hum Genet 55(2):266–277
Wille H, Requena JR (2018) The structure of PrPSc prions. Pathogens 7(1):20
Young JC, Hoogenraad NJ, Hartl FU (2003) Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. Cell 112(1):41–50

Management of Insulin Through Co-Solute Engineering: A Therapeutic Approach

12

Sania Bashir, Neha Sami, Sayema Bashir, Faizan Ahmad, Md. Imtaiyaz Hassan, and Asimul Islam

Abstract

Fibrillation and aggregation of protein are some of the most exciting frontiers in protein chemistry and molecular medicine. It is also expected to shed light on the molecular and biochemical basis of various pathological conditions having a dramatic social impact such as Alzheimer's, Parkinson's diseases, and type II diabetes. The role of insulin in different physiological processes, effect on its synthesis and secretion, along with its actions on the molecular level to the whole-body level, has important implications in chronic diseases prevailing in westernized populations today. Rapid globalization, urbanization, and industrialization have spawned epidemics of obesity, diabetes, and their attendant comorbidities, like physical inactivity and dietary imbalance, unmask latent predisposing genetic traits. The present review discusses insulin, its structure, and its etiology in diabetes. The aggregation mechanism begins with the diffusion of insulin and then its adsorption at the hydrophobic interface that leads to conformational changes of oligomers to expose the hydrophobic residues. Factors influencing its aggregation, such as temperature, light exposure, pH, salt, protein concentration, drying, and agitation, have also been discussed. At last, the therapeutic approaches with recent drug interventions have also been mentioned.

S. Bashir · F. Ahmad · M. I. Hassan · A. Islam (✉)
Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
e-mail: aislam@jmi.ac.in

N. Sami
Department of Biosciences, Jamia Millia Islamia, New Delhi, India

S. Bashir
Faculty of Dentistry, Jamia Millia Islamia, New Delhi, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_12

Keywords

Insulin · Fibrils · Osmolytes · Therapeutics · Neurodegenerative diseases · Diabetes aggregation · Protein misfolding · Clinical drugs

12.1 Introduction

Obesity, diabetes, and Alzheimer's disease (AD) are among the most classy and disabling disorders globally. The correlation between cognitive impairment and metabolic diseases has remained undetected. But nowadays, many epidemiological pieces of evidence explain the secure link between these conditions. The most significant marker of metabolic dysregulation, i.e., insulin resistance, is a factor also present in AD (McEvoy et al. 2012; Yaffe et al. 2006). In this chapter, we will discuss insulin, its structure, insulin resistance, and its etiology in different disorders. This chapter will also provide a vision for the improvement and progression for various formulations of insulin and improve the bioavailability of insulin powder formulation that would give an alternate treatment option having better adequacy or tolerability among the patients when contrasted with the intravenous conveyance. This would offer a better way to administer epidemic diabetes.

12.2 Insulin

Insulin assumes a central job in the guideline of human metabolism. The hormone is a 51-buildup anabolic protein that is secreted by the β-cells in the Islets of Langerhans. Containing two chains (A and B) associated with disulfide bonds, the developed hormone is the posttranslational product of a single chain precursor, assigned proinsulin (Fig. 12.1). Broad investigations of the three-dimensional structure of insulin, pioneered by D. C. Hodgkin, have empowered the improvement of helpful analogs to treat the metabolic disorder diabetes mellitus (Vestergaard et al. 2007). The insulin gene is the site of dominant mutation transformations related to diabetes mellitus (Hua and Weiss 2004; Vestergaard et al. 2007).

12.2.1 Structure of Insulin

Human insulin is a globular pre-protein with 110 amino acid with a molecular weight of 11980.91 Da. The active protein is a 51-residue hormone of two chains, chain A has 21 residues while chain B has 30 residues (Table 12.1).

These chains are linked by two disulfide bridges, which join the A-chain's N- and C-terminal helices to the B-chain central helix (Fig. 12.2). In proinsulin, C-terminus is connected to A-chain and B-chain by N-terminus peptide links (Ahmad et al. 2004; Blundell et al. 1972; Wilcox 2005). It is having an isoelectric point at pH 5.5 (DeFronzo 2010; Wilcox 2005) and theoretical pI 5.22 (uniport).

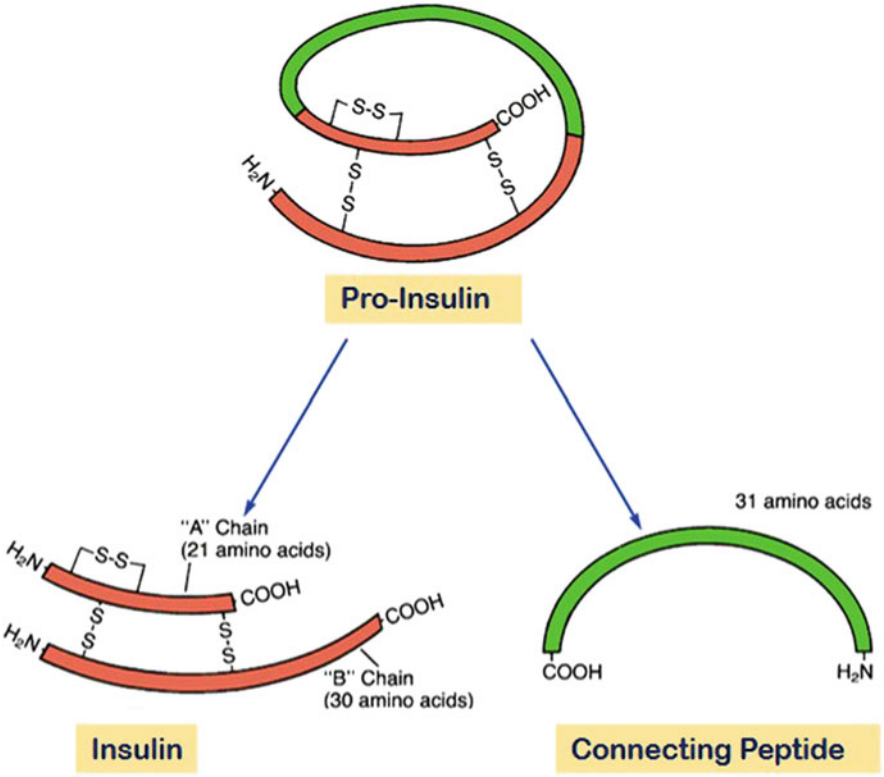


Fig. 12.1 Reorganization from proinsulin to insulin. Pancreas secretes the immediate precursor, which is proinsulin. Removal of the C-peptide generates active insulin which comprises two peptide chains as chain A and chain B and two disulfide bonds (S–S) which are connected. Proteases remove associating peptide (C-peptide) to discharge dynamic insulin, made out of two peptide chains (A and B) associated by two disulfide (S–S) bonds. Since C-peptide emerges just from endogenous insulin, its quality in blood shows that at least some pancreatic insulin is being made

Insulin is present at deficient concentrations in the bloodstream. At these low concentrations, it exists in a monomeric form, which is the biologically active form. However, it forms dimers and hexamers at higher concentrations and in the presence of zinc ions, respectively, at neutral pH. Each of the two zinc ions binds to three His-B10 residues, which is very important for the hexamer formation. In insulin crystals, up to ten additional zinc ions bind in other sites. The hexamer exists in three different conformations: T6, T3R3, and R6, depending on the conformations of the monomer subunits. The principal difference being that residues B1–B8 are converted from an extended conformation in the T-state to an R-helix in the R-state. Ligands also play an important role in changing the conformation from the T- to the R-state by binding to the allosteric sites on the hexamer (Chang et al. 1997).

12.2.1.1 Insulin Monomer

The structure of monomer arrangement has a view of perpendicular threefold crystal axis, with two interchain disulfide bonds A7-B7 and A20-B19 connecting the end of

Table 12.1 Insulin chain A and chain B amino acid sequence

Chain A	Amino acid	Side chain polarity	Side chain charge	Hydrophobicity	Chain B	Amino acid	Side chain polarity	Side chain charge	Hydrophobicity
A_1	Gly	Nonpolar	Neutral	−0.4	B_1	Phe	Nonpolar	Neutral	2.8
A_2	Ile	Nonpolar	Neutral	4.5	B_2	Val	Nonpolar	Neutral	
A_3	Val	Nonpolar	Neutral	4.2	B_3	Asn	Polar	Neutral	−3.5
A_4	Glu	Polar	Negative	−3.5	B_4	Gln	Polar	Neutral	−3.5
A_5	Gln	Polar	Neutral	−3.5	B_5	His	Polar	Positive	−3.2
A_6	Cys	Nonpolar	Neutral	2.5	B_6	Leu	Nonpolar	Neutral	3.8
A_7	Cys	Nonpolar	Neutral	2.5	B_7	Cys	Nonpolar	Neutral	2.5
A_8	Thr	Nonpolar	Neutral	−0.7	B_8	Gly	Nonpolar	Neutral	−0.4
A_9	Ser	Polar	Neutral	−0.8	B_9	Ser	Polar	Neutral	−0.8
A_{10}	Ile	Nonpolar	Neutral	4.5	B_{10}	His	Polar	Positive	−3.2
A_{11}	Cys	Nonpolar	Neutral	2.5	B_{11}	Leu	Nonpolar	Neutral	3.8
A_{12}	Ser	Polar	Neutral	−0.8	B_{12}	Val	Nonpolar	Neutral	4.2
A_{13}	Leu	Nonpolar	Neutral	3.8	B_{13}	Glu	Polar	Negative	−3.5
A_{14}	Tyr	Polar	Neutral		B_{14}	Ala	Nonpolar	Neutral	1.8
A_{15}	Gln	Polar	Neutral	−3.5	B_{15}	Leu	Nonpolar	Neutral	3.8
A_{16}	Leu	Nonpolar	Neutral	3.8	B_{16}	Tyr	Polar	Neutral	−1.3
A_{17}	Glu	Polar	Negative	−3.5	B_{17}	Leu	Nonpolar	Neutral	3.8
A_{18}	Asn	Polar	Neutral	−3.5	B_{18}	Val	Nonpolar	Neutral	4.2
A_{19}	Tyr	Polar	Neutral	−1.3	B_{19}	Cys	Nonpolar	Neutral	2.5
A_{20}	Cys	Nonpolar	Neutral	2.5	B_{20}	Gly	Nonpolar	Neutral	−0.4
A_{21}	Asn	Polar	Neutral	−3.5	B_{21}	Glu	Polar	Negative	−3.5
					B_{22}	Arg	Polar	Positive	−4.5
					B_{23}	Gly	Nonpolar	Neutral	−0.4
					B_{24}	Phe	Nonpolar	Neutral	2.8
					B_{25}	Phe	Nonpolar	Neutral	2.8

				B_{26}	Tyr	Polar	Neutral	−1.3
				B_{27}	Thr	Nonpolar	Neutral	−0.7
				B_{28}	Pro	Nonpolar	Neutral	−1.6
				B_{29}	Lys	Polar	Positive	−3.9
				B_{30}	Thr	Nonpolar	Neutral	−0.7

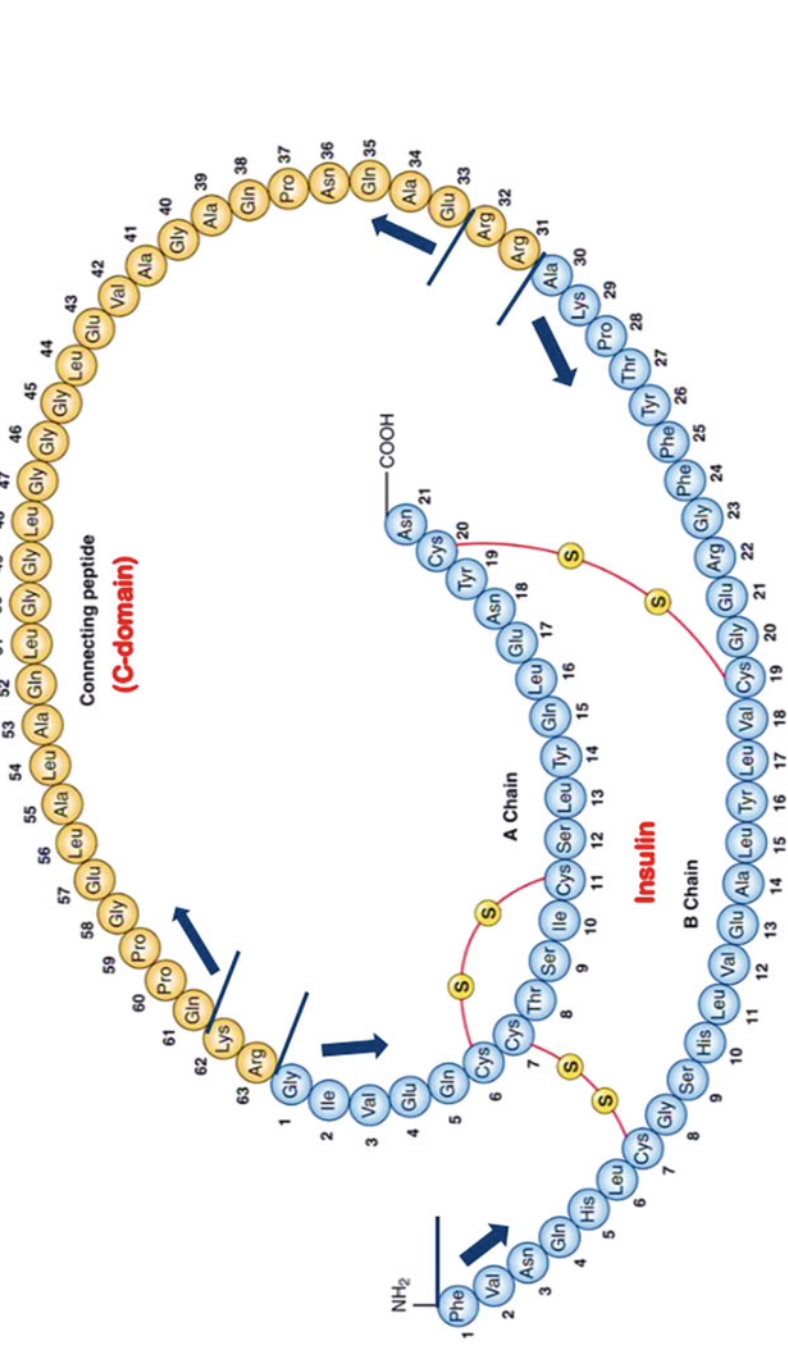


Fig. 12.2 Conversion of proinsulin to insulin showing the secondary and tertiary structures of insulin; the flexible C domain in proinsulin and split proinsulins is shown in the dotted line. Thick blue arrows indicate the absolute path

its central α-helix from chain A to chain B that makes the backbone. A-chain has A6 and A11 has an internal disulfide bond between them. The C-terminal residues that bring back the N-terminals beneath form the second helical structure (not α-helix) at A13, which folds back across the face of the molecule. The initial stretches of the helix are both cystines A6 and A7. The surface molecules are residues of A7 connecting to B7, A8, A9, & A10. The B-chain above the U-turn takes the arrangement of the final three A-chains at B20–B23. The chains A and B are in tight arrangements because of H-bond in between carbonyl oxygen at A19 and α-nitrogen at B25 of this region. The compactly organized complete molecule contains the free B-chain in the monomer structure. The completely buried non-polar residues including cystines (A6-A9 and A20-B19), leucines (A16, B11 and B15), and the isoleucine (A2) provide the insulin monomer with a hydrophobic core that plays an important role in its stabilization in the aqueous media. The valine of B12 and tyrosine of B16 make up large nonpolar surfaces as the C-terminal B-chain residues remain on the surface of the molecule. Another nonpolar region comprises alanine of B14, leucine of B17, and valine of B18 of the B-chain α-helix, and phenylalanine of B1 at the N-terminus and the nearby two A-chain residues leucine of A13 and tyrosine of A14 (Weiss et al. 2014).

12.2.1.2 Insulin Dimer

The dimers molecule that corresponds to asymmetric crystal units is related by the local axis, which is not exactly seen. The main dimer contacts are between nonpolar side chains making up the surface molecule, due to such arrangements the α-helix of chain B is brought together up with the C-terminal residues, which runs anti-parallel to each other causing the twofold relationship between the molecules (Fig. 12.3). This makes possible the formation of anti-parallel β-pleated sheet structure containing four hydrogen bonds between the monomers (Yao et al. 1999). The two among the four hydrogen bonds are between B12 of one molecule and B26 of the other; the remaining two are between equivalent residues across the local axis. The anti-parallel β-pleated sheet is the only secondary structure that shows twofold symmetry and bridges twofold related molecules. The valines of B12 and phenylalanines of B24 are buried in the dimer, whereas the phenylalanines of B25 are at the surface involving the equivalent close hydrophobic contacts along the local axis. These hydrogen bonds of the β-pleated sheet are partially covered, and the two phenylalanines of B25 collapsed together on surface forming dimer, possibly aiding their structural role (Sowdhamini et al. 1989). The stability of the solution is due to hydrogen bonds having nonpolar contacts explaining the directional properties present in inward of the dimer. The dimer is formed when either the B-chain of α-helix or the β-pleated sheet is formed. Thus, it forms a good balance between nonpolar and hydrogen binding forces (Baker et al. 1988; Blundell et al. 1972).

12.2.1.3 Insulin Hexamer

Hexamer structure of insulin forms because of zinc ions' presence at neutral pH . The aggregate of three dimers together in a threefold symmetrical way forms a hexamer (Fig. 12.4). The central cause of hexamer formation is because of the binding of two

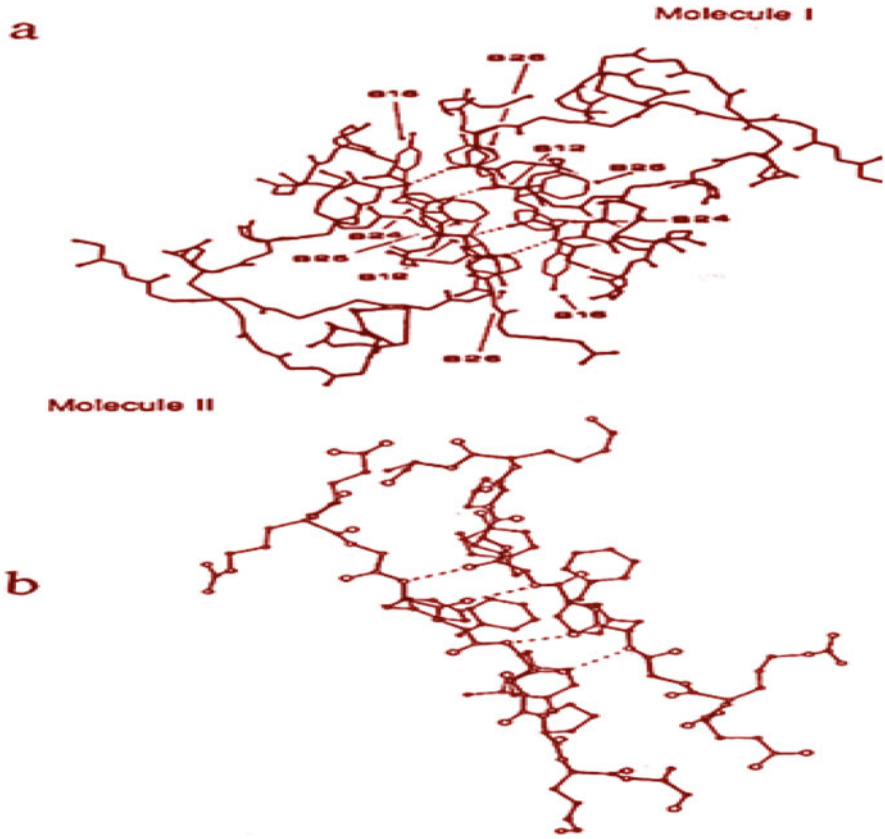


Fig. 12.3 Structural illustration of the monomer–monomer interface in the insulin dimer. (**a**) The dimer is viewed along the crystallographic twofold axis. The side chains of residues Val ^{B12}, Leu ^{B15}, Phe ^{B24}, Phe ^{B25}, and Tyr ^{B26}, which form the core of the insulin dimer, are illustrated in the figures (not labeled: Pro ^{B28}). Four main-chain hydrogen bonds are formed from the main-chain atoms of Phe ^{B24}, and Tyr ^{B26} are illustrated as dotted lines. (**b**) Magnified view of the dimer interface showing the four main-chain hydrogen bonds.

zinc ions, which is about 18 Å apart lying on the threefold axis and proximately related by the local axis. The crystal uses the threefold hexamer axis of the rhombohedral. The three N3 imidazole nitrogens are co-ordinated by each zinc ions having threefold B10 equivalent histidines. Binding of zinc to the hexamer structure by the dimers organization results in both polar and nonpolar interactions between the molecules. Nonpolar residues are present at the envelope of axis relating to the dimer. The B-chain α-helix has the important central residues of alanine B14, leucine B17, and valine B14.

The organization of the dimers into the zinc-binding hexamer structure leads to both polar and nonpolar interactions between the molecules. There are nonpolar

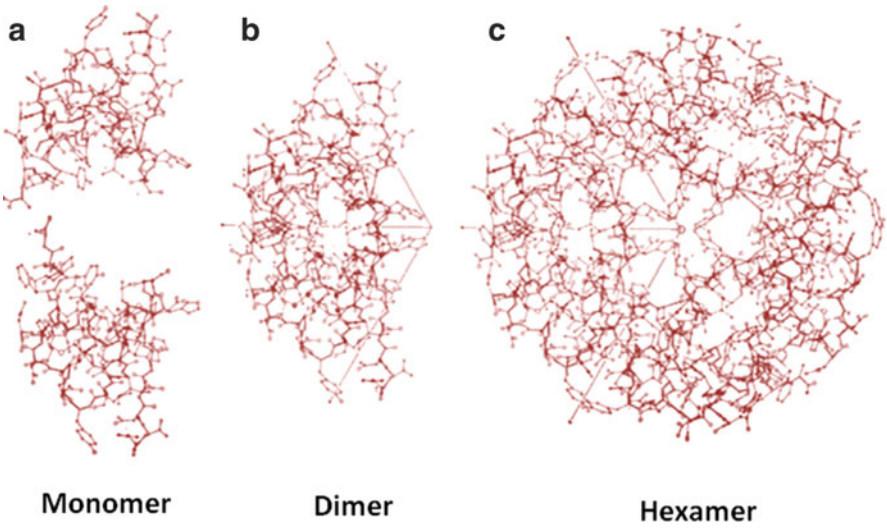


Fig. 12.4 The relationship between the (**a**) Insulin monomers, (**b**) Insulin monomers bond to form the dimer, and (**c**) Three insulin dimers molecule aggregate to form the hexamer. The direction of the local axis relating to the monomer is indicated at each stage of aggregation by the arrow

residues along with the envelope of the axis relating the dimers, and the essential central residues involved are the B-chain α-helix residues B14 alanine, B17 leucine, and B18 valine. In the center of the hexamer, down the threefold axis is a hydrophilic core made up of residues B9 serine, B10 histidine, and B13 glutamic acid and their threefold-related equivalents. These polar groups are connected by a water molecule (Smith et al. 2003). The dimer organization in the hexamer brings the positively charged N-terminal amino groups close to the negatively charged A17 glutamic acid. In one case, the two groups appear to be developing a well-defined salt bridge structure. The hexamer has the state of a compressed disc, around 50 Å across and 35 Å high. The focal point of the hexamer has a hydrophilic center consisting of serine B9, histidine B10, and glutamic acid B13. A detailed water structure associates these polar gatherings.

12.3 Protein Aggregation and Insulin

Protein aggregation occurs in vivo due to improper folding. Many diseases emerge from protein misfolding and are precisely assembled under "protein conformational diseases," including the vast majority of the neurodegenerative disorders, for example, Parkinson's disease (PD), AD, Huntington's disease (HD), prion encephalopathies (PrP^{Sc}), sickle cell anemia, cystic fibrosis, and diabetes. The hallmark event of all these diseases is the structural change in the protein's secondary or tertiary structure, leading to the formation of protein aggregates having

different supra-molecular organizations (Fig. 12.5). These aggregates then form amyloid deposits that are organized in structurally well-defined fibrils. The primary feature of the amyloidogenic proteins is their structural instability induced either by posttranslational alterations, mutations, or by local conditions, similar to temperature, pH, and co-solutes. The conformational change advances diseases either by the increase of toxic action or by the absence of biological function of the natively folded protein. The characterization of these aggregates needs to be determined to target therapeutic approaches for these diseases.

12.3.1 Insulin Aggregation/Fibrillation Mechanisms

The kinetic process of insulin aggregation is characterized by the lag phase followed by an exponential growth phase and a final plateau/stationary phase. The apparent lag phase is due to nucleation process which is reversible, the oligomers (insulin aggregates) are undetectable by present methods. Earlier studies suggest that the nucleation of insulin results from a simultaneous assemblage of few misfolded insulin monomers into oligomers through their hydrophobic surfaces (Nielsen et al. 2001c; Sluzky et al. 1992). The exponential growth phase occurs when nucleation concentration reaches a critical state. Thus further addition of insulin monomers or oligomers to nuclear leads to long unbranched fibrils (Nielsen et al.

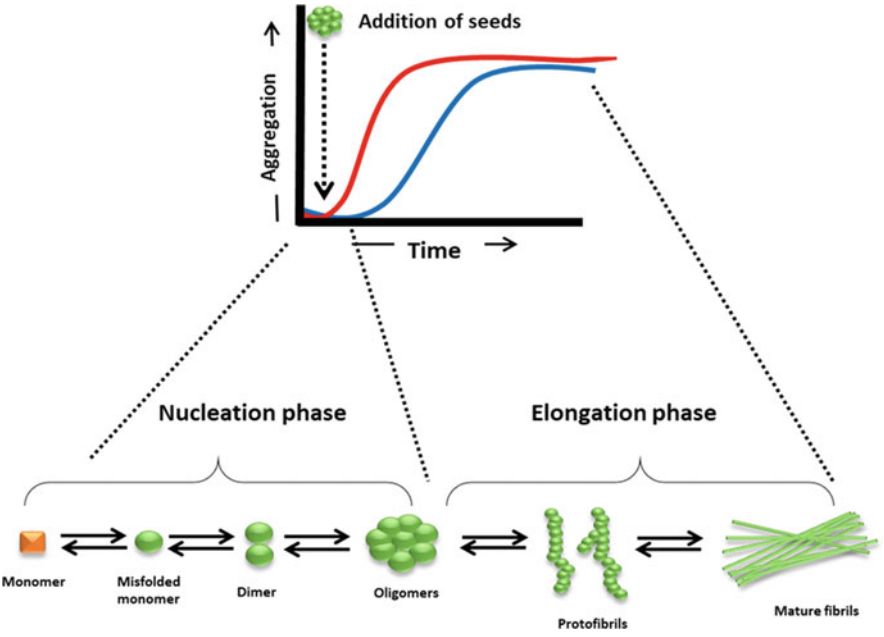


Fig. 12.5 Idealized schematic representation of a general overall pathway for the formation of protein fibrils, including some possible intermediates

2001c). A stable plateau state is reached when the concentration of insulin in solution reaches below the threshold, thus restricts fibrils extension (Fig. 12.5). The changing monomeric insulin to oligomers and to develop mature fibrils during the process of insulin fibrillation are believed to be (1) hydrophobic residues limitations to aqueous condition, (2) hydrogen bonding saturation, and (3) global free-energy minimum of non-native alternative state. Partially folded intermediate is formed initially as a result of misfolding monomer. As a result, the hydrophobic residues become exposed to a solvent which is normally covered in the dimer and hexamer (Brange et al. 1997; Nielsen et al. 2001c). When two monomers with conformational change associate together, a mutual anti-parallel β-sheet structure is formed, which is not the same as normal insulin dimer. This distinctive structure in the assemblage makes the monomer–monomer bond essentially more potent than the association in the dimer and hexamer (Mollmann et al. 2005). The consequent advances of association of insulin molecules to the nucleic may result from an effort to limit the hydrophobic residuals' exposure. This process advances rich fibril formation.

Amyloid development consists of two stages: (1) a nucleation phase or lag phase, in which monomers experience conformational change/misfolding and associate to form oligomeric cores, and (2) an elongation phase or growth phase, in which the nuclei rapidly develop by further accumulation of monomers and bigger polymers until saturation. Sigmoidal curve represents the kinetics of amyloid formation by lag phase followed by growth phase (blue curve), where aggregation determines the rate-limiting step during formation of seed or nuclei. The rate restricting advance in the process is the arrangement of cores/seeds to advance total. Along these lines, amyloid arrangement can be significantly speed up by the expansion of preformed seeds (cores). The accumulation of seeds decreases the lag time and induces earlier aggregate formation (red curve). They form cross-β structures through the formation of intermolecular and intermolecular hydrogen bonding (Jansen et al. 2005).

12.4 Characterization of Protein Aggregation

12.4.1 Protein Solubility

The change in protein solubility, whether comparing various ionic strengths, heat treatments, or combinations, is critical (Melander and Horváth 1977; Shaw et al. 2001). Solubility has been evaluated based on centrifugation for a range of g-forces and time, then measuring soluble (supernatant) and insoluble (pellet) protein. The protein solubility can be calculated as a percent of protein content in the supernatant versus total protein content by centrifuging the homogenized solutions at 20000 × g for 15 min. Throughout the literature, investigators have used an assortment of both g-forces and time to report solubility of various protein systems. Ideal centrifugation parameters would be those which simulate a beverage system's shelf life. The correlation between centrifugation parameters and shelf life of a protein beverage has not been published in the literature.

12.4.2 Turbidity Measurements

Turbidity can be measured using turbidimeters or spectrophotometers. Turbidity measurements using spectrophotometers and turbidity meters are nondestructive, rapid, inexpensive, and require small sample sizes (Hu et al. 2017; Shaw et al. 2001). The light from a tungsten lamp emits polychromatic or white light. The emitted light is then separated into individual wavelengths by a prism so that the aperture can select a specific wavelength. The light enters the sample chamber where it is either scattered in multiple directions, transmitted, and/or absorbed. A photocell detects the transmitted light, and the percent of absorbance or transmittance can be provided. Optical density can be monitored at various wavelengths from 400 to 700 nm. 400 nm is often used for protein solutions since this wavelength is sensitive to particulate size while not being absorbed. Particles scatter light waves in a multitude of directions, but a turbidimeter evaluates only that which is at a 90° angle between its photo-detector and incident light beam. A negative bias, where the measurement observed is lower than actual turbidity, occurs when the particle is not physically dense. A disadvantage occurs with pigmentation since light could be absorbed, decreasing the amount of scattered light to the detector. Molecular movement is the source of the instability of turbidity. This can be minimized by controlled temperatures and time frame of evaluation (Chantrapornchai et al. 2001). The Orbeco-Hellige turbidimeter uses a tungsten filament lamp that is associated with a large bandwidth of wavelengths, which is beneficial for evaluating a wide range of particle sizes. Interferences contribute to a greater extent when evaluating samples with high turbidity.

12.5 Techniques Used to Study Insulin Aggregation

Kinetic and structural studies have been done to decipher the molecular mechanism of insulin aggregation or fibrillation (Ahmad et al. 2005; Jansen et al. 2005; Nielsen et al. 2001c; Smith et al. 2007; Sorenson and Drummond 2014). Insulin fibrils bind with the dyes like Thioflavin T (ThT) and Congo red to emit characteristic fluorescence; this has been demonstrated through the kinetic studies (Khurana et al. 2001; Nielsen et al. 2001a, c; Smith et al. 2007). Spectroscopic methods include Fourier transform infrared (FTIR) and circular dichroism (CD) (Table 12.2). These have been used to study the conformational changes in protein during fibrillation (Brange et al. 1997; Nielsen et al. 2001b, c). Limitation of FTIR and CD is that they indicate only the secondary structural changes (i.e., α-helix, β-sheets, and random coil), and thus provide limited structural information studies. Electron microscopy (EM) and Atomic force microscopy (AFM) have been used to directly visualize the insulin morphologies during the fibrillation pathway (Jimenez et al. 2002; Manno et al. 2006; Nielsen et al. 2001b). They reveal the insulin fibril architecture as unbranched, long, and straight with periodic twists but not the atomic-level resolution.

Emerging techniques attempt to elucidate the atomic-level structures of protein in 3D are small-angle X-ray scattering, X-ray micro crystallography, and solid-state

Table 12.2 Techniques to study aggregation

UV	CD	Fluorescence	ITC	DLS	DSC	FTIR	TEM	ANS	ThT	CR	SEC	MS	X-ray	Computational	References
	✔	✔	✔				✔		✔					✔	
✔	✔				✔			✔	✔						Saha and Deep (2016)
✔	✔	✔					✔		✔						
	✔			✔			✔		✔		✔				Gong et al. (2015)
✔	✔	✔	✔												Choudhary et al. (2015)
✔	✔						✔		✔		✔				Ahmad et al. (2004)
✔	✔				✔										Huus et al. (2005)
	✔	✔					✔					✔			
✔	✔	✔		✔											Kumar et al. (2013)
	✔								✔		✔				
			✔		✔	✔									Dzwolak et al. (2003)
	✔					✔	✔						✔		Bouchard et al. (2000)

nuclear magnetic resonance (NMR) spectroscopy (Brange et al. 1997; Jimenez et al. 2002; Vestergaard et al. 2007). The structure of visualization methods is being challenged by the insoluble and uncrystallizable nature of insulin fibrils. Nucleation mechanism determines the fate of aggregation of insulin by partially folded intermediates (Gazit 2002; Hua and Weiss 2004; Sluzky et al. 1992). Finally, most of the work done to understand insulin aggregation has been summarized in Table 12.2. The first row of Table 12.1 shows the possible techniques to follow insulin aggregation, while the last column of the table provides the references for the readers. The tick in the columns (1–15) donates that the technique has been used in the paper stated in column 16. This table provides an easy search for the students who want to study insulin aggregation as various techniques and their references are provided in detail.

12.6 Factors Influencing Protein Aggregation In Vitro

Various factors influence the aggregation in proteins. Some major factors are shown in Table 12.3.

12.6.1 Temperature

Temperature plays a major factor in changing the structure and functions of the protein, distorting their three-dimensional structure. The stability that is maintained due to bonds of covalent and non-covalent is distorted due to external factors like temperature. The unfolded to folded states are because of energy barriers. Thus, the native protein ought to be in more stable conformation than the unfolded (Chaudhuri et al. 2014). Proteins maintain their structure, functions, and stability by sustaining their optimum temperature. The instability due to exposure of heat and conformation leads to the denaturation of proteins since the exposed hydrophobic regions are substantially more, these exposed hydrophobic regions lead to aggregation formulation (Wang 1999; Wang et al. 2010). Protein unfolding can also be induced in lowering the temperature, which is reported to enhance the oligomerization or cluster formation in monoclonal antibodies (Esfandiary et al. 2015; Fukuda et al. 2017; Godfrin et al. 2013; Raut and Kalonia 2015; Salinas et al. 2010). Protein–protein interactions are of the major attribute at low temperatures. Though it is not clear for high temperature-sensitive, what kind of key interactions are responsible (Salinas et al. 2010).

12.6.2 Exposure to Light

The photolytic and non-photolytic degradation of proteins is induced due to light exposure. Thus the products formed due to degradation may have monomers or aggregates which enhance the aggregation tendency. Antibody solutions in UV or

Table 12.3 Factors affecting aggregation of proteins in vivo

Category	Factors	Structure	pH and buffer	Effects on protein and aggregation	References
Physical factors	Temperature	Native α helical	2.3	Partially/completely unfolding of protein. Protein–protein interaction or chemical degradations, leading to aggregation or crosslink proteins directly	Bouchard et al. (2000)
		On heating to 70 °C			
		Above 60 °C			Brange et al. (1997)
		Acidic solu. within few hours at 37 °C			
		Aqueous solution at 37 °C	7.0		
		Zn-free, unfolded at 70 °C		Induces the adsorption of surface protein, influencing the stability, promoting the protein aggregation by the partial unfolding of the protein.	Huus et al. (2005)
		68 °C for 2 h	2.6		
		68 °C	2.4		
	Light/ irradiation				
	Container/ closure systems				
Chemical factors	Buffering agent and concentration	Urea Gdn HCl NaCl	2 7 7.4	Alters colloidal stability. Displaying variable effects, such as ion–protein interaction and specificity or the ionic effect. Protein–ion interactions, balancing the general charge-screening effect and interaction of protein–protein interference. High concentration or additives tends to aggregate normally.	Ahmad et al. (2004); Shukla et al. (2017)
	pH				
	Ionic strength				
	Excipients (additives)				
Processing factors	Purification			Depending on the variables of host system, expressions and fermentation conditions such as freeze-thawing	
	Refolding				
	Freeze-thawing				

(continued)

Table 12.3 (continued)

Category	Factors	Structure	pH and buffer	Effects on protein and aggregation	References
	Shaking			promote protein–ice interactions, leading to unfolding and adsorption. Air-liquid protein adsorption induced by shaking results unfolding. Shearing exposes hydrophobic patches, pasteurization have minimal effect at low or moderate pressures, but at over 100 Mpa, it induces structure loss. Drying loss of hydrational layers leads to disruption of the structure of protein enhancing aggregation of protein.	
	Pressurization				
	Fermentation/ expression				
	Drying				
	Shearing				
	Analytical methodologies				

visible light exposed higher molecular weight in some cases (Mason et al. 2012; Singh et al. 2012; Sreedhara et al. 2016). Light exposure generates singlet oxygen as reactive oxygen species (ROS). These light exposures generate a significant amount of tryptophan (Trp) and methionine (Met) as an oxidized product which enhances aggregation (Agarkhed et al. 2013; Singh et al. 2012; Sreedhara et al. 2016).

12.6.3 pH

Sensitivity induced in protein due to structure and function deviation is a major reason for pH change, any unconventionality in pH results in a distortion of protein structure (Kameoka et al. 2007). On the fact creating dipole because of positive and negative charge in a protein creates a charge of attraction or repulsion. These zwitterions or charged ions provide an attraction force with opposite charges forming oligomers, leading to the pathway of aggregates. The isoelectric point plays a crucial role in the formation of such dipoles. On the contrary, the pH of the media with acidic or basic affects the surface charged protein if they have deviated from their isoelectric point. Thus the formation of aggregates takes place due to electrostatically unfavorable conditions by self-association of molecules (Chi et al. 2003; Katayama et al. 2006). The stability of the protein weather colloidal or conformational is instantly reliant on the solution condition determining pH influencing aggregation (Galm et al. 2017). Protein–protein interactions can be altered due to shift in the pH solution or protein affect their cross-linking leading to nucleation and growth of aggregates (Brummitt et al. 2011; Calero-Rubio et al.

2018; Esfandiary et al. 2015; Sahin et al. 2012; Zhang et al. 2017). A higher rate of aggregation of protein was noted in interferon-tau when used in phosphate buffer while slower compared to histidine and Tris-buffers (Roefs and De Kruif 1994).

12.6.4 Protein Concentration

The concentration of protein determines one of the crucial factors for its aggregation. Generally, at higher concentrations, protein tends to aggregate by incubating or storing for a longer duration (Alford et al. 2008; Fields et al. 1992). The concept of macromolecular crowing develops from such a situation where the molecules of the protein are more than the volume of medium, creating the condition of macromolecular crowding. These crowders are reported as the stimulant for protein aggregation (Minton 2005, 2008). β-Lactoglobulin produces aggregates at a higher concentration which are relatively larger in size and number (Fields et al. 1992; Roefs and De Kruif 1994).

12.6.5 Salt

Salts are the common obstruct to the electrostatic interactions, which ultimately affect the structure, stability, and function of a protein. Since the protein core is a hydrophobic residue by which surface is enclosed via typically hydrophilic residue, it maintains its integrity and typical structure through electrostatic interactions (Desai et al. 2017). Thus these charged amino acids due to hydrophilic interactions interact with solvents in solutions. Where salt at high concentration induces aggregation, low salt concentration betters the solubility of the protein. The reasons are obvious as by the time salts at higher concentrations forming salt ions in the solvent molecule are being solubilized, protein finds its way to interact with other molecules of protein, leading to the pathway of generating aggregates (Chi et al. 2003; Matthew 1985).

12.6.6 Surfactants

Amphipathic molecules having both hydrophobic and hydrophilic groups attached are commonly known as surfactants. These molecules may take part in stabilization, destabilization, or induction of protein aggregation (Wang et al. 2017b). The surfactant molecules as anions and cation neutralize protein surface charges, which in results partially unfold the protein structure, favoring agglomeration of protein via hydrophobic interactions (Chaturvedi et al. 2016; Khan et al. 2012; Siddiqi et al. 2016). Both hydrophobic amino acid residue and tail of hydrophobic surfactant be responsible for both hydrophobicities (Khan et al. 2012). Ionic strength affects the conformational stability. The effect of ionic strength influences leads to protein–protein interactions because of the conformational stability and nature of the protein (Arzensek et al. 2015; Bickel et al. 2016; Calero-Rubio et al. 2018).

12.6.7 Macromolecular Crowding

The macromolecular crowding environment helps protein to attain its three-dimensional structure by changing its kinetic properties and thermodynamics. The stress caused due to molecular crowding or crowding agents leads to fibril aggregates, which results in several types of diseases as reported (Munishkina et al. 2004). Types of crowders like PEG (polyethylene glycol), ficoll 70, dextran that are hydrophilic are spherical polymers. Ficoll 70 and PEG 3.5 are reported to enhance fibrillation in α-lactalbumin and α-synuclein proteins, and human insulin in monomeric form (Munishkina et al. 2004; Shahid et al. 2017). Their extent of acceleration depends on the concentration and nature of a crowder. They work on the principle of the surface-to-volume ratio, which is commonly known as the excluded volume effect. There is some inverse effect of crowding agents, such as with lysozyme, where the initial addition of crowding agent in their native form inhibits the fibrillation of protein (Munishkina et al. 2008).

12.6.8 Agitation

The process of agitation includes shaker, rotator, vortex, and stirrer or any form of mixing (Bai et al. 2012). Shaking or stirring is one of the common processing factors for inducing aggregation in peptides or proteins (Dengl et al. 2013; Gandhi et al. 2017; Lewis et al. 2017; Shah et al. 2017; Wang et al. 2017a). During agitation-induced protein unfolding/aggregation or shearing stress, there is a development of air–liquid interface which leads to the exposure of hydrophobic surface (Zhai et al. 2012) due to stirring bar or simple grinding and abrasion (Lin et al. 2016; Nesta et al. 2017; Sediq et al. 2016). Agitation-induced aggregates may differ morphologically due to presence or absence of covalent bonds like disulfide bonds depending on their stress condition (Brych et al. 2010; Telikepalli et al. 2014). Thus partial dissociation due to non-covalent interactions may reverse the formation into monomers with time (Dekel et al. 2017; Kiese et al. 2010).

12.6.9 Freeze-Thawing

Freezing has prompted aggregation in many proteins (Tokhadze et al. 2018), which is due to many reasons such as structural changes or perturbation-induced denaturation due to low temperature (Hauptmann et al. 2018; Paul et al. 2017; Vlieland et al. 2018). Freeze-thawing promotes unfolding and adsorption due to protein and ice interactions (Twomey et al. 2013). Freeze-thawing affects aggregation of the protein as the freezing rate can influence aggregation or stability of protein during storage (Desai et al. 2017; Mezhebovsky et al. 2016).

12.6.10 Drying

Protein hydration layer potential removed due to the several drying processes is also one of the factors in promoting aggregation (Kumar et al. 2009). Spray-drying or freeze-drying process forms small intensities of aggregates by conventional freeze-drying (Faghihi et al. 2014; Zhou et al. 2016). Meanwhile, is not clear that drying, per se, was the cause.

12.7 Diabetes: A Metabolic Syndrome

Diabetes is a genuine and interminable issue that can be ascribed to the lacking arrival of the insulin or when the body does not react to the insulin, which is as of now present. This illness has been focused on therapeutic activity and execution among four nontransferable infections by world pioneers, as shown in a WHO report. The commonness of diabetes has been evaluated to be twofold from 4.7% of the grown-up population in 1980 to 8.5% in the present situation. The estimated deaths due to this epidemic disease were 1.5 million in 2012 alone (Collaboration ERF 2010). Currently, approximately 400 million people are suffering from diabetes worldwide. Although the variables provoking Type I diabetes are obscure, there are many treatments accessible for Type II illness, which packs over 90% of the complete diabetes cases. Be that as it may, the subside and flow treatment for this sickness is by all accounts deficient as far as appropriate administration of disease, while insulin despite everything remains an ultimate treatment to accomplish similarly successful glycemic control (Tandon et al. 2018).

12.8 Osmolytes

Osmolytes are naturally occurring small organic molecules (Khan et al. 2010), with lower molecular weight helps in preserving the characteristics of the organism by maintaining the biological fluid. Osmolytes thus maintain the integrity of the solution for their ionic property, viscosity, and melting point. These osmolytes affect the strength of proteins and nucleic acids in aqueous solutions (Singh et al. 2011; Yancey 2005; Yancey and Siebenaller 2015), and enhance the strength and stability of protein devoid of hampering their activity (Yancey et al. 1982). Osmolytes are classified into various categories, mentioning in sub-categorical units as polyols and sugars, such as monosaccharides like glucose, fructose, disaccharides like sucrose, trehalose, and polyols such as glycerol and sorbitol. Amino acids and derivatives like proline, taurine, and glycine (Fig. 12.6).

Some osmolytes, particularly affecting insulin protein, have been mentioned in Table 12.4. Methylammonium salt as betaine, glycerophosphoryl choline (GPC), tri-methylamine N-oxide (TMAO), and sarcosine hence are chosen as protective and stabilizing osmolytes for proteins (Yancey 2005; Yancey and Siebenaller 2015). Some osmolytes like urea, however, destabilizes the protein. Other defensive

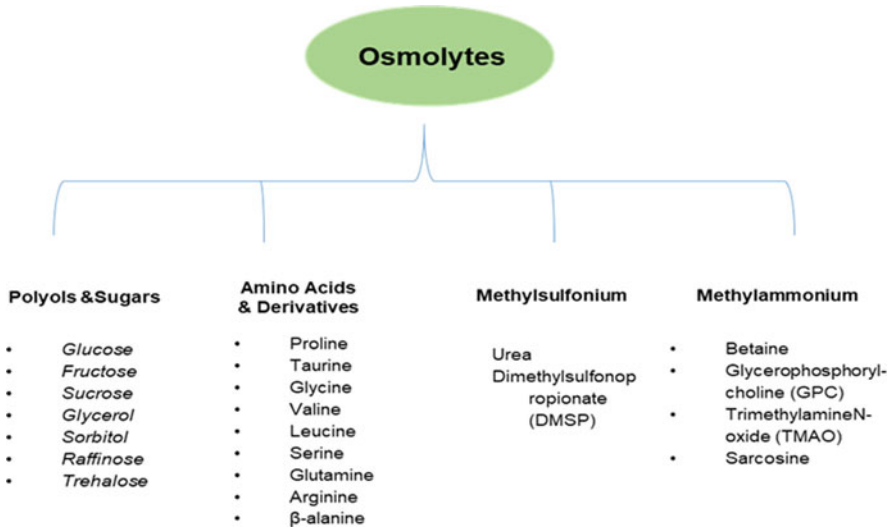


Fig. 12.6 Classification of osmolytes

osmolytes like trehalose, proline, betaine, and TMAO also destabilizes the protein but at higher concentration and at non-physiological pH by effecting the protein-specific interactions.(Singh et al. 2011; Yancey and Siebenaller 2015).

Protein stability depends on the theory of osmophobicity, surface tension, preferential exclusion principle that is protein binding to water, excluded volume effect (Beg et al. 2015) and water replacement theory which propose the maintenance of native protein conformation by replacing H-bond between water and protein which in results stabilizes the protein (Naik et al. 2006; Wlodarczyk et al. 2018). Based on diffusion, the peptide backbone and side-chain formed due to free energy transmission lead to the interaction of peptide backbones functional group to osmolytes, combating their hostile interaction for stabilizing the protein (Chandel et al. 2017). Amino acids, such as arginine and lysine, are common names for resolving fibrils and inclusion bodies. High-stress (HS) fibrils due to absence osmolytes "shielding" it from the high-stress conditions, low stress (LS) are the resulting fibrils here. The resulting fibrils are here called low stress (LS) (Bajorunaite et al. 2007; Das et al. 2007; Macchi et al. 2012) (Fig. 12.7).

Therefore, mentioning osmolytes as virtuous folders along with suitable stabilizers would not be incorrect (Singh et al. 2011). The fantastic work of osmolytes is not limited till stability but also in assisting the correct folding of the misfolded proteins, by increasing their intercellular function and preventing their degradation (Eleutherio et al. 1998; Russo et al. 2003; Singh et al. 2007; Tatzelt et al. 1996; Welch and Brown 1996). Prior studies proposed that misfolding of gene aquaporin-2 (AQP2) protein in mammals leads to diabetes insipidus. Conversely when1M glycerol was added to the cell culture medium, it induced proper

Table 12.4 Effect of osmolytes addition to insulin fibrillating proteins

Osmolytes	Effect on insulin	References
Glucose	Minor effect	Nayak et al. (2009)
Sucrose	No/minor effect	Nayak et al. (2009); Arora et al. (2004)
Fructose	Minor effect	Nayak et al. (2009)
Sorbitol	Reduced fibrillation/ low rate fibril	Choudhary et al. (2015)
Glycerol	Low rate fibril	
Trehalose	Reduced/slower fibrillation	Nayak et al. (2009); Arora et al. (2004)
Proline	Highly reduced fibrillation	Choudhary et al. (2015)
Taurine		
Ectoine	Highly reduced fibrillation	Arora et al. (2004)
Betaine	Highly reduced fibrillation	Arora et al. (2004); Choudhary et al. (2015)
Citrulline	Reduced fibrillation	Choudhary et al. (2015)
TMAO	Low rate fibril	Choudhary et al. (2015)
Quinones Benzoquinones (BQ)	Inhibit amyloid aggregation	Gong et al. (2015)
Napthoquinones (NQ)		
Anthroquinones (AQ)	Strongly inhibit β-amyloid protein	Gong et al. (2015)
Phenanthroquinones (PQ)		
Protic ionic liquid PILs: (TMAS, TEAHS, TMAP, TEAP, TMAA)	Reduced in presence of PILs	
Curcumin	Amyloid β-sheet	
EGCG	Reduced in presence of PILs	Kumar et al. (2013)
Resveratrol	Amyloid β-sheet	Alam et al. (2017)

rearrangement and reshuffling such that the protein folds to its native form (Khan et al. 2019; Tamarappoo et al. 1999).

Many neurological and metabolic disorders in pathophysiology conditions is a dramatic social impact such as AD, PD, HD, PrP^{Sc}, Dementia, Amyotrophic lateral sclerosis (ALS), familial amyotrophic lateral sclerosis (FALS), cystic fibrosis, type-2 diabetes, and particular forms of emphysema (Stefani 2004). Studies due reported the beneficiary effects of these chaperons and a potential suppressor of fibril causing neurodegenerative diseases because of useful therapeutic targets to retain into their native conformation by avoiding disorders or fibril formation (Inayathullah and Rajadas 2016). Studies have publicized the avert effect on fibrils and aggregates of protein in the presence of sarcosine, L-proline, 4-hydroxy-L-proline, and TMAO (Choudhary and Kishore 2014). Proline with three molars and above concentration has reported averting the accumulation of bovine carbonic anhydrase (BCA) (Mittal

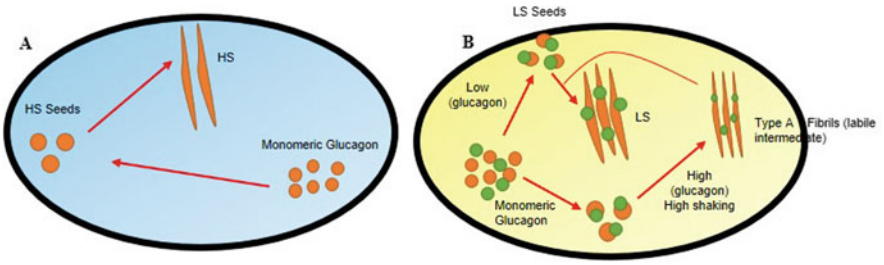


Fig. 12.7 Fibrillation mechanism. (**a**) Absence of osmolytes leads to high-stress (HS) fibrils. (**b**) Presence of Pro, Ser, or COS leads to the formation of the labile type A intermediate as the osmolytes bind glucagon, "shielding" it from the high-stress conditions. The resulting fibrils are called low stress (LS)

and Singh 2014). This marked proline in another story of publication as a "defensive agent in inhibition of protein aggregation" due to the capacity of sinking the abnormal interactions between polypeptide chain at the early onset of the trail in pathogenic cause due to protein aggregation (Ignatova and Gierasch 2006). An investigation endeavored to examine the role of osmolytes in the amyloid-coupled collection setup on the human model insulin hormone protein. The result witnessed that sorbitol, glycerol, betaine, quinones, and TMAO observed a low rate of fibrillation formation due to the progression of reducing the rate of unfolded monomers. The formerly mentioned investigational results have shown a fantastic connection through volume isolation rule applicable to polymer crowding (Khan et al. 2019).

12.9 Therapeutic Approaches

12.9.1 Gene Therapy

Gene therapy is an intracellular delivery of transgene (genomic materials) into specific cells to generate a therapeutic impact by correcting the existing prevailing abnormality from the norm and providing the cells with novel function (Stone 2010). Many gene delivery systems possibly are applied in gene therapy to turning off any particular gene or restore a specific gene function. A definitive objective of gene therapy is single management of a suitable material to replace a defective or missing gene. Presently, popular methods, for example, lentivirus, adenovirus, and AAV just as nonviral strategies, like liposomes and naked DNA, have been utilized for insulin gene conveyance into various tissues, for instance, pancreas, liver, adipocytes, and muscle (Wang et al. 2004; Wang et al. 2007; Zhang et al. 2001). Remarkably, intestinal cells, for example, enteroendocrine K-cells, demonstrated numerous similitudes with pancreatic β-cells, it produces glucose-dependent insulinotropic polypeptide (GIP) and contains prohormone convertases vital for proinsulin formation (Cho and Kieffer 2010). Therefore, researches are being done

for in vitro modification of K-cells to create and discharge insulin, but the implantation of these cells has failed to reverse diabetes.

12.9.2 Pharmacological Drug Approach

When contrasted with the way of life intercessions alone, each pharmacological operator used as monotherapy multiplied the number of patients who came to HbA1c target levels underneath 7% by two to three folds. Because of poor diabetes control, most patients would require different treatments to accomplish great glycemic control over the long route (Nathan et al. 2009). Hypoglycemia is one of the major risk factors that should be paid special attention to when assortments of various pharmacological agents are utilized associatively. The decision of which pharmacological specialists to be picked for every patient ought to be a standard dynamic procedure. The cost, potential reactions, potential advantages, glucose bringing down viability, and dosing routine are things to be thought about before choosing a prescription. Dose modifications are essential for renal disabled patients. Regular observation is fundamental for renal hindered patients; however, for each one of those on pharmacological agents (White 2014).

12.9.3 Insulin Injection

It remains the primary treatment for T1-DM, where insulin lack is seen. At the point when oral hypoglycemic medications are not effective in regularizing glucose and HbA1c levels in T2-DM, insulin can be used as monotherapy or together with oral hypoglycemic specialists. The constraining element of insulin is that it must be directed through infusions. Despite showing favorable treatment effects, needle fear causes weak consistency, which prompts inadequate glycemic control. Insulin pumps can be considered as continuous subcutaneous insulin infusions, which are accessible in the market now (McCall and Farhy 2013). The US Food and Drug Administration (FDA) has characterized these pumps as moderate to high hazard gadgets, and clinical preliminaries are consistently being accomplished for security and adequacy purposes (Sorenson and Drummond 2014). A nonintrusive option would be through breathed in or oral insulin. Challenges are set up to utilize these courses of organizations. Improved pharmacokinetic and pharmacodynamic parameters may guarantee the achievement of insulin using this organization course. Oral insulin is another engaging chance. It is still in the stage of clinical trials, and a lot more research is in store before this formulation hits market shelves (Table 12.5).

Table 12.5 Ongoing clinical trials for diabetes (www.ClinicalTrials.gov)

Clinical trial number	Drugs	Phase	Status
NCT02692313	Epinephrine, saline infusion	Phase 1	Recruiting
NCT00953914	Pyridostigmine, placebo	NA	Completed
NCT02641548	Sock of silicone, heel cream	NA	Recruiting
NCT00682903	Guardian R	NA	Completed
NCT01030770	Ranibizumab, 0.9% sodium chloride	Phase 3	Completed
NCT00482768	Practice facilitation	NA	Completed
NCT03757845	Mediterranean diet	NA	Recruiting
NCT02835287	Protocol based integrated care	NA	Not recruiting
NCT02473809	Liraglutide, placebo	Phase 4	Completed
NCT00703989	Benfotiamine, lipoic acid	NA	Completed
NCT02505451	Dobutamine	NA	Completed
NCT01554644	Prontosan, saline	NA	Withdrawn
NCT03877523	Cocarnit	NA	Completed
NCT02541838	Muscle, power, balance, perturbation, aerobic exercise	NA	Completed
NCT03562208	Adult bone marrow stem cells	Phase 3	Recruiting
NCT03497260	Fructose, plain water	NA	Completed
NCT01695278	Telephone counseling	Phase 3	Completed
NCT00701181	Laser treatment, PF-04523655 high, PF-04523655 middle, PF-04523655 low	Phase 2	Terminated
NCT03289338	Zoledronic acid, methylprednisolone, placebos	Phase 2	Completed
NCT01823406	Euglycemic clamp (normal blood sugar clamp) for 4 h hyperglycemic clamp (elevated blood sugar to 300) for 4 h	NA	Completed
NCT01821508	Clinical treatment, roux-en-Y gastric bypass surgery	NA	Not recruiting
NCT00318214	MRE0094, vehicle gel	Phase 2	Terminated
NCT00312364	MRE0094	Phase 2	Completed
NCT03908762	iSage app with connected glucometer magnet	NA	Recruiting
NCT00272831	Cilostazol, placebo	Phase 4	Completed
NCT00815178	Inspiratory muscle training, inspiratory muscle training placebo	NA	Completed
NCT03392441	Insulin deprivation in type 1 diabetic patients	NA	Not recruiting
NCT03530176	18 fluoride-sodium fluoride radioisotope	NA	Completed

(continued)

Table 12.5 (continued)

Clinical trial number	Drugs	Phase	Status
NCT02915263	IGIV-C, 0.9% sodium chloride	Phase 2	Recruiting
NCT02631902	Exercise, dietary intervention	NA	Completed
NCT01926522	Technological rehabilitation, control rehabilitation	NA	Completed
NCT00279266	Thiamine	NA	Withdrawn
NCT03437421	Vitamin D3 (cholecalciferol)	NA	Recruiting
NCT03378271	Simultaneous CGM/FGM	NA	Recruiting
NCT04238208	Lidocaine 5% patch, capsaicin 8% patch, per 10^2 cm	Phase 4	Completed
NCT03208309	Diacerein, placebo	Phase 2	Completed
NCT03639545	Empagliflozin 25 mg, metformin empagliflozin/ metformin, placebos	Phase 4	Recruiting
NCT01571297	RM-131, placebo	Phase 2	Completed
NCT01371318	Online wound electronic medical record	NA	Completed
NCT03906383	Remote ischemic conditioning	NA	Completed
NCT02169167	Resin salve treatment, octenidine treatment	NA	Completed
NCT00337298	Amlodipine, lisinopril	NA	Completed
NCT01394055	RM-131, placebo	NA	Completed
NCT01454700	Insulin pump therapy (CSII) plus continuous glucose monitoring (CGM), multiple daily insulin injections (MDI)	Phase 4	Completed
NCT00539409	Pulsatile intravenous insulin therapy (humulin R, Novolog)	Phase 3	Terminated
NCT03878277	Starbucks® cold brew—325 mL bottle	Phase 2	Enrolling by invitation
NCT03899883	Pegloticase 8 mg/mL [krystexxa]	Phase 2	Not yet recruiting
NCT03618420	Aminohippurate, sodium inj 20%, iohexol inj 300 mg/mL	Phase 2	Active, not recruiting
NCT04058951	High animal protein diet (HAPD), high plant protein diet (HPPD)	NA	Recruiting
NCT03620773	Aminohippurate sodium inj 20%, iohexol inj 300 mg/mL, vertical sleeve gastrectomy	Phase 2	Recruiting
NCT02155361	Topical Citrullus colocynthis, fruit oil (1%), topical vehicle oil	Phase 2	Completed
NCT01746563	Ranibizumab, laser therapy	Phase 2	Completed
NCT02704494	Resveratrol, losartan, placebo	Phase 1	Completed
NCT04224428	Fexofenadine pill, placebo oral tablet	Phase 4	Recruiting

(continued)

Table 12.5 (continued)

Clinical trial number	Drugs	Phase	Status
NCT04170543	MEDI3506, placebo	Phase 2	Recruiting
NCT04125329	Human umbilical cord mesenchymal stem cells	Phase 1	Not yet recruiting
NCT04026165	SEL, placebo	Phase 3	Recruiting
NCT03933956	Empagliflozin 10 MG	Phase 3	Recruiting
NCT03869931	Fenofibrate	Phase 3	Recruiting
NCT03824379	Magnesium citrate, antidiabetic	Phase 2	Recruiting
NCT03804879	LMB763, placebo	Phase 2	Recruiting
NCT03165227	BI 685509, placebo	Phase 1	Completed

12.10 Conclusion

The aggregation of insulin starts with dispersion and adsorption of insulin at the hydrophobic solid, resulting in conformational changes of the monomeric insulin to uncover the hydrophobic buildups. The driving force is due to hydrophobic interaction between the expanded monomer (unfolded) and the hydrophobic interface. Hostile conditions like temperature, pH, protein concentration, salt, surfactants, molecular crowders, agitation, thawing, drying, as discussed in the review, prompt the generation of the misfolded aggregated structure of the protein that further prompts the age of the neurological issue. Osmolytes contribute to combatting the aggregation and protein misfolding. They can be used as remedial applications for numerous neurological issues that arise due to protein misfolding, which are principally connected with the protein misfolding. Fibrillation of protein is obligated for various amyloidogenic issues inspecting diseases like AD, HD, PD, cystic fibrosis, diabetes type 2, and dialysis connected amyloidosis. A detailed comprehension of the mechanisms of action of osmolytes can prompt the extension of osmolytes as an ancient remedial target molecule and thus to steady drug design for the avoidance and cure of neurological, hereditary, and other ailments brought about by protein misfolding/fibrillation/conglomeration alongside different other factors. Since the molecular foundation of aggregation is varied, increasingly model frameworks ready to recognize those pathways will be expected to get a clearer picture.

Additionally, more significant arrangements of osmolytes just as model proteins should be screened to reach determinations on their general mechanism of action,

due to the explicitness of some osmolyte–protein cooperation. Although some advancement has been made lately on the aggregation of insulin at the interfaces, an enormous number of open questions despite everything exist about the fundamental changes and pathways during the procedure of insulin aggregation. Some of those questions can be as (a) a molecular-level comprehension of the accumulation pathway from native insulin to insulin oligomers and developed mature fibril strands is as yet deficient and lacking; (b) the structure and misfolding pathway of briefly populated oligomeric insulin species stay to be identified and elucidated. Varied strategies used to ponder such rational, determined disease to have various degrees of stage trail drugs are pleasing and engaging redirection for the social being.

References

Agarkhed M, O'Dell C, Hsieh MC, Zhang J, Goldstein J, Srivastava A (2013) Effect of polysorbate 80 concentration on thermal and photostability of a monoclonal antibody. AAPS PharmSciTech 14(1):1–9

Ahmad A, Millett IS, Doniach S, Uversky VN, Fink AL (2004) Stimulation of insulin fibrillation by urea-induced intermediates. J Biol Chem 279(15):14999–15013

Ahmad A, Uversky VN, Hong D, Fink AL (2005) Early events in the fibrillation of monomeric insulin. J Biol Chem 280:42669–42675

Alam P, Siddiqi K, Chturvedi SK, Khan RH (2017) Protein aggregation: from background to inhibition strategies. Int J Biol Macromol 103:208–219

Alford JR, Kendrick BS, Carpenter JF, Randolph TW (2008) High concentration formulations of recombinant human interleukin-1 receptor antagonist: II. Aggregation kinetics. J Pharm Sci 97:3005–3021

Arora A, Ha C, Park CB (2004) Inhibition of insulin amyloid formation by small stress molecules. FEBS Lett 564(1–2):121–125

Arzensek D, Kuzman D, Podgornik R (2015) Hofmeister effects in monoclonal antibody solution interactions. J Phys Chem B 119(33):10375–10389

Bai G, Bee JS, Biddlecombe JG, Chen Q, Leach WT (2012) Computational fluid dynamics (CFD) insights into agitation stress methods in biopharmaceutical development. Int J Pharm 423:264–280

Bajorunaite E, Sereikaite J, Bumelis VA (2007) L-arginine suppresses aggregation of recombinant growth hormones in refolding process from E. coli inclusion bodies. Proteins 26:547–555

Baker EN, Blundell TL, Cutfield JF, Dodson EJ, Dodson GG, Hodgkin DMC, Hubbard RE, Isaacs NW, Reynolds CD, Sakabe K (1988) The structure of 2Zn pig insulin crystals at 1.5 Å resolution. Philos Trans R Soc Lond B Biol Sci 319:369–456

Beg I, Minton AP, Hassan MI, Islam A, Ahmad F (2015) Thermal stabilization of proteins by mono-and oligosaccharides: measurement and analysis in the context of an excluded volume model. Biochemistry 54:3594–3603

Bickel F, Herold EM, Signes A, Romeijn S, Jiskoot W, Kiefer H (2016) Reversible NaCl-induced aggregation of a monoclonal antibody at low pH: characterization of aggregates and factors affecting aggregation. Eur J Pharm Biopharm 107:310–320

Blundell T, Cutfield J, Cutfield S, Dodson E, Dodson G, Hodgkin D, Mercola D (1972) Three-dimensional atomic structure of insulin and its relationship to activity. J Diabetes 21:492–505

Bouchard M, Zurdo J, Nettleton EJ, Dobson CM, Robinson CV (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. Protein Sci 9(10):1960–1967

Brange J, Dodson G, Edwards D, Holden P, Whittingham J (1997) A model of insulin fibrils derived from the X-ray crystal structure of a monomeric insulin (despentapeptide insulin). Proteins 27:507–516

Brummitt RK, Nesta DP, Chang L, Chase SF, Laue TM, Roberts CJ (2011) Nonnative aggregation of an IgG1 antibody in acidic conditions: part unfolding, colloidal interactions, and formation of high-molecular-weight aggregates. J Pharm Sci 100:2087–2103

Brych SR, Gokarn YR, Hultgen H, Stevenson RJ, Rajan R, Matsumura M (2010) Characterization of antibody aggregation: role of buried, unpaired cysteines in particle formation. J Pharm Sci 99:764–781

Calero-Rubio C, Ghosh R, Saluja A, Roberts CJ (2018) Predicting protein-protein interactions of concentrated antibody solutions using dilute solution data and coarse-grained molecular models. J Pharm Sci 107:1269–1281

Chandel TI, Khan MV, Khan RH (2017) Impact of osmolytes in conformational modulation of protein and its applications in biotechnology. In: Singh LR, Dar TA (eds) Cellular osmolytes: from chaperoning protein folding to clinical perspectives. Springer, Singapore, pp 143–160

Chang X, Jørgensen AMM, Bardrum P, Led JJ (1997) Solution structures of the R6 human insulin hexamer. Biochemistry 36:9409–9422

Chantrapornchai W, Clydesdale F, McClements D (2001) Influence of flocculation on optical properties of emulsions. J Food Sci 66:464–469

Chaturvedi SK, Khan JM, Siddiqi MK, Alam P, Khan RH (2016) Comparative insight into surfactants mediated amyloidogenesis of lysozyme. Int J Biol Macromol 83:315–325

Chaudhuri R, Cheng Y, Middaugh CR, Volkin DB (2014) High-throughput biophysical analysis of protein therapeutics to examine interrelationships between aggregate formation and conformational stability. AAPS J 16:48–64

Chi EY, Krishnan S, Randolph TW, Carpenter JF (2003) Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. Pharm Res 20:1325–1336

Cho YM, Kieffer TJ (2010) K-cells and glucose-dependent insulinotropic polypeptide in health and disease. Vitam Horm 84:111–150

Choudhary S, Kishore N (2014) Addressing mechanism of fibrillization/ aggregation and its prevention in presence of osmolytes: spectroscopic and calorimetric approach. PLoS One 9 (8):e104600

Choudhary S, Kishore N, Hosur RV (2015) Inhibition of insulin fibrillation by osmolytes: mechanistic insights. Sci Rep 5:17599

Collaboration ERF (2010) Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 375:2215–2222

Das U, Hariprasad G, Ethayathulla AS, Manral P, Das TK, Pasha S, Mann A, Ganguli M, Verma AK, Bhat R (2007) Inhibition of protein aggregation: supramolecular assemblies of arginine hold the key. PLoS One 2(11):e1176

DeFronzo RA (2010) Insulin resistance, lipotoxicity, type 2 diabetes and atherosclerosis: the missing links. The Claude Bernard Lecture 2009. Diabetologia 53:1270–1287

Dekel Y, Machluf Y, Gefen T, Eidelshtein G, Kotlyar A, Bram Y, Shahar E, Reslane F, Aizenshtein E, Pitcovski J (2017) Formation of multimeric antibodies for self-delivery of active monomers. Drug Deliv 24:199–208

Dengl S, Wehmer M, Hesse F, Lipsmeier F, Popp O, Lang K (2013) Aggregation and chemical modification of monoclonal antibodies under upstream processing conditions. Pharm Res 30:1380–1399

Desai KG, Pruett WA, Martin PJ, Colandene JD, Nesta DP (2017) Impact of manufacturing-scale freeze-thaw conditions on a mAb solution. Bio Pharm Int 30:30–36

Dzwolak W, Ravindra R, Lendermann J, Winter R (2003) Aggregation of bovine insulin probed by DSC/PPC calorimetry and FTIR spectroscopy. Biochemistry 42(38):11347–11355

Eleutherio EC, Silva JT, Panek AD (1998) Identification of an integral membrane 80 kDa protein of Saccharomyces cerevisiae induced in response to dehydration. Cell Stress Chaperones 3:37
Esfandiary R, Parupudi A, CasasFinet J, Gadre D, Sathish H (2015) Mechanism of reversible self-association of a monoclonal antibody: role of electrostatic and hydrophobic interactions. J Pharm Sci 104:577–586
Faghihi H, Vatanara A, Najafabadi AR, Ramezani V, Gilani K (2014) The use of amino acids to prepare physically and conformationally stable spray-dried IgG with enhanced aerosol performance. Int J Pharm 466:163–171
Fields GB, Alonso DO, Stigter D, Dill KA (1992) Theory for the aggregation of proteins and copolymers. J Phys Chem A 96:3974–3981
Fukuda M, Watanabe A, Hayasaka A, Muraoka M, Hori Y, Yamazaki T, Imaeda Y, Koga A (2017) Small-scale screening method for low-viscosity antibody solutions using small-angle X-ray scattering. Eur J Pharm Biopharm 112:132–137
Galm L, Amrhein S, Hubbuch J (2017) Predictive approach for protein aggregation: correlation of protein surface characteristics and conformational flexibility to protein aggregation propensity. Biotechnol Bioeng 114:1170–1183
Gandhi AV, Pothecary MR, Bain DL, Carpenter JF (2017) Some lessons learned from a comparison between sedimentation velocity analytical ultracentrifugation and size exclusion chromatography to characterize and quantify protein aggregates. J Pharm Sci 106:2178–2186
Gazit E (2002) The "correctly folded" state of proteins: is it a metastable state? Angew Chem Int Ed Engl 41:257–259
Godfrin PD, CastañedaPriego R, Liu Y, Wagner NJ (2013) Intermediate range order and structure in colloidal dispersions with competing interactions. J Chem Phys 139:154904
Gong H, He Z, Peng A, Zhang X, Cheng B, Sun Y, Zheng L, Huang K (2015) Effects of several quinones on insulin aggregation. Sci Rep 4(1):5648
Hauptmann A, Podgoršek K, Kuzman D, Srčič S, Hoelzl G, Loerting T (2018) Impact of buffer, protein concentration and sucrose addition on the aggregation and particle formation during freezing and thawing. Pharm Res 35(5):101
Hu YT, Ting Y, Hu JY, Hsieh SC (2017) Techniques and methods to study functional characteristics of emulsion systems. J Food Drug Anal 25:16–26
Hua Q, Weiss MA (2004) Mechanism of insulin fibrillation the structure of insulin under amyloidogenic conditions resembles a protein-folding intermediate. J Biol Chem 279:21449–21460
Huus K, Havelund S, Olsen HB, van de Weert M, Frokjaer S (2005) Thermal dissociation and unfolding of insulin. Biochemistry 44(33):11171–11177
Ignatova Z, Gierasch LM (2006) Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. PNAS 103:13357–13361
Inayathullah M, Rajadas J (2016) Effect of osmolytes on the conformation and aggregation of some amyloid peptides: CD spectroscopic data. Data Brief 7:1643–1651
Jansen R, Dzwolak W, Winter R (2005) Amyloidogenic self-assembly of insulin aggregates probed by high resolution atomic force microscopy. Biophys J 88:1344–1353
Jimenez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, Saibil HR (2002) The protofilament structure of insulin amyloid fibrils. PNAS 99:9196–9201
Kameoka D, Masuzaki E, Ueda T, Imoto T (2007) Effect of buffer species on the unfolding and the aggregation of humanized IgG. J Biochem 142:383–391
Katayama DS, Nayar R, Chou DK, Valente JJ, Cooper J, Henry CS, VanderVelde DG, Villarete L, Liu C, Manning MC (2006) Effect of buffer species on the thermally induced aggregation of interferon-tau. J Pharm Sci 95:1212–1226
Khan SH, Ahmad N, Ahmad F, Kumar R (2010) Naturally occurring organic osmolytes: from cell physiology to disease prevention. IUBMB Life 62:891–895
Khan JM, Qadeer A, Chaturvedi SK, Ahmad E, Rehman SAA, Gourinath S, Khan RH (2012) SDS can be utilized as an amyloid inducer: a case study on diverse proteins. PLoS One 7(1):e29694

Khan S, Mueed Z, Deval R, Rai PK, Prajapati DK, Poddar NK (2019) Role of osmolytes in amyloidosis. In: Surguchov A (ed) Synucleins-biochemistry and role in diseases. IntechOpen, London

Khurana R, Uversky VN, Nielsen L, Fink AL (2001) Is Congo red an amyloid-specific dye? J Biol Chem 276:22715–22721

Kiese S, Pappenberger A, Friess W, Mahler HC (2010) Equilibrium studies of protein aggregates and homogeneous nucleation in protein formulation. J Pharm Sci 99:632–644

Kumar V, Sharma VK, Kalonia DS (2009) In situ precipitation and vacuum drying of interferon alpha-2a: development of a single-step process for obtaining dry, stable protein formulation. Int J Pharm 366:88–98

Kumar A, Venkatesu P (2013) Prevention of insulin self-aggregation by a protic ionic liquid. RSC Adv 3(2):362–367

Lewis LM, Pizzo ME, Sinha S, Ahmed SS, Joseph L (2017) Visible and sub-visible particle formation for a model bioconjugate. AAPS PharmSciTech 18:926–931

Lin GL, Pathak JA, Kim DH, Carlson M, Riguero V, Kim YJ, Buff JS, Fuller GG (2016) Interfacial dilatational deformation accelerates particle formation in monoclonal antibody solutions. Soft Matter 12:3293–3302

Macchi F, Eisenkolb M, Kiefer H, Otzen DE (2012) The effect of osmolytes on protein fibrillation. Int J Mol Sci 13:3801–3819

Manno M, Craparo EF, Martorana V, Bulone D, San Biagio PL (2006) Kinetics of insulin aggregation: disentanglement of amyloid fibrillation from large-sizecluster formation. Biophys J 90:4585–4591

Mason D, Schneich C, Kerwin A (2012) Effect of pH and light on aggregation and conformation of an IgG1 mAb. Mol Pharm 9:774–790

Matthew JB (1985) Electrostatic effects in proteins. Annu Rev Biophys 14:387–417

McCall A, Farhy L (2013) Treating type 1 diabetes: from strategies for insulin delivery to dual hormonal control. Minerva Endocrinol 38(2):145

McEvoy LK, Laughlin GA, BarrettConnor E, Bergstrom J, KritzSilverstein D, Der-Martirosian C, von Mühlen D (2012) Metabolic syndrome and 16-year cognitive decline in community-dwelling older adults. Ann Epidemiol 22:310–317

Melander W, Horváth C (1977) Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. Arch Biochem Biophys 183:200–215

Mezhebovsky T, Routhier E, Sass P, Shahrokh Z (2016) Enabling freeze-thaw stability of PBS-based formulations of a monoclonal antibody. BioPharm Int 29:33–39

Minton AP (2005) Influence of macromolecular crowding upon the stability and state of association of proteins: predictions and observations. J Pharm Sci 94:1668–1675

Minton AP (2008) Effective hard particle model for the osmotic pressure of highly concentrated binary protein solutions. Biophys J 94:L57–L59

Mittal S, Singh LR (2014) Macromolecular crowding decelerates aggregation of a β-rich protein, bovine carbonic anhydrase: a case study. J Biochem 156:273–282

Mollmann S, Bukrinsky J, Frokjaer S, Elofsson U (2005) Adsorption of human insulin and AspB28 insulin on a PTFE-like surface. J Colloid Interface Sci 286:28–35

Munishkina LA, Henriques J, Uversky VN, Fink AL (2004) Role of protein−water interactions and electrostatics in α-synuclein fibril formation. Biochemistry 43:3289–3300

Munishkina LA, Ahmad A, Fink AL, Uversky VN (2008) Guiding protein aggregation with macromolecular crowding. Biochemistry 47:8993–9006

Naik V, Kardani J, Roy I (2006) Trehalose-induced structural transition accelerates aggregation of α-synuclein. Mol Biotechnol 58:251–255

Nathan DM, Buse JB, Davidson MB, Ferrannini E, Holman RR, Sherwin R, Zinman B (2009) Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: ADA. Diabetes Care 32:193–203

Nayak A, Lee C-C, McRae GJ, Belfort G (2009) Osmolyte controlled fibrillation kinetics of insulin: new insight into fibrillation using the preferential exclusion principle. Biotechnol Prog 25 (5):1508–1514

Nesta D, Nanda T, He J, Haas M, Shpungin S, Rusanov I, Sweder R, Brisbane C (2017) Aggregation from shear stress and surface interaction: molecule-specific or universal phenomenon? Bioprocess Int 15(4):30–39

Nielsen L, Frokjaer S, Brange J, Uversky VN, Fink AL (2001a) Probing the mechanism of insulin fibril formation with insulin mutants. Biochemistry 40:8397–8409

Nielsen L, Frokjaer S, Carpenter JF, Brange J (2001b) Studies of the structure of insulin fibrils by Fourier transform infrared (FTIR) spectroscopy and electron microscopy. J Pharm Sci 90:29–37

Nielsen L, Khurana R, Coats A, Frokjaer S, Brange J, Vyas S, Uversky VN, Fink AL (2001c) Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. Biochemistry 40:6036–6046

Paul AJ, Bickel F, Röhm M, Hospach L, Halder B, Rettich N, Handrick R, Herold EM, Kiefer H, Hesse F (2017) High-throughput analysis of sub-visible mAb aggregate particles using automated fluorescence microscopy imaging. Anal Bioanal Chem 409:4149–4156

Raut AS, Kalonia DS (2015) Opalescence in monoclonal antibody solutions and its correlation with intermolecular interactions in dilute and concentrated solutions. J Pharm Sci 104:1263–1274

Roefs SP, De Kruif KG (1994) A model for the denaturation and aggregation of β-lactoglobulin. Eur J Biochem 226:883–889

Russo AT, Rösgen J, Bolen D (2003) Osmolyte effects on kinetics of FKBP12 C22A folding coupled with prolyl isomerization. J Mol Biol 330:851–866

Saha S, Deep S (2016) Glycerol inhibits the primary pathways and transforms the secondary pathway of insulin aggregation. Phys Chem Chem Phys 18(28):18934–18948

Sahin E, Weiss WF IV, Kroetsch AM, King KR, Kessler RK, Das TK, Roberts CJ (2012) Aggregation and pH–temperature phase behavior for aggregates of an IgG2 antibody. J Pharm Sci 101:1678–1687

Salinas BA, Sathish HA, Bishop SM, Harn N, Carpenter JF, Randolph TW (2010) Understanding and modulating opalescence and viscosity in a monoclonal antibody formulation. J Pharm Sci 99:82–93

Sediq AS, van Duijvenvoorde RB, Jiskoot W, Nejadnik MR (2016) No touching! Abrasion of adsorbed protein is the root cause of subvisible particle formation during stirring. J Pharm Sci 105:519–529

Shah M, Rattray Z, Day K, Uddin S, Curtis R, van der Walle CF (2017) Pluen evaluation of aggregate and silicone-oil counts in pre-filled siliconized syringes: an orthogonal study characterising the entire subvisible size range. Int J Pharm 519:58–66

Shahid S, Hassan MI, Islam A, Ahmad F (2017) Size-dependent studies of macromolecular crowding on the thermodynamic stability, structure and functional activity of proteins: in vitro and in silico approaches. Biochim Biophys Acta 1861:178–197

Shaw KL, Grimsley GR, Yakovlev GI, Makarov AA, Pace CN (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. Protein Sci 10:1206–1215

Shukla H, Kumar R, Sonkar A, Mitra K, Akhtar MS, Tripathi T (2017) Salt-regulated reversible fibrillation of *Mycobacterium tuberculosis* isocitrate lyase: concurrent restoration of structure and activity. Int J Biol Macromol 104:89–96

Siddiqi MK, Shahein YE, Hussein N, Khan RH (2016) Effect of surfactants on Ra-sHSPI–A small heat shock protein from the cattle tick Rhipicephalus annulatus. J Mol Struct 1119:12–17

Singh LR, Chen X, Kožich V, Kruger WD (2007) Chemical chaperone rescue of mutant human cystathionine β-synthase. Mol Genet Metab 91:335–342

Singh LR, Poddar NK, Dar TA, Kumar R, Ahmad F (2011) Protein and DNA destabilization by osmolytes: the other side of the coin. Life Sci 88:117–125

Singh SR, Zhang J, O'Dell C, Hsieh MC, Goldstein J, Liu J, Srivastava (2012) Effect of polysorbate 80 quality on photostability of a monoclonal antibody. AAPS PharmSciTech 13:422–430

Sluzky V, Klibanov AM, Langer R (1992) Mechanism of insulin aggregation and stabilization in agitated aqueous solutions. Biotechnol Bioeng 40:895–903
Smith GD, Pangborn WA, Blessing RH (2003) The structure of T6 human insulin at 1.0 Å resolution. Acta Crystallogr D Biol Crystallogr 59:474–482
Smith M, Sharp J, Roberts C (2007) Nucleation and growth of insulin fibrils in bulk solution and at hydrophobic polystyrene surfaces. Biophys J 93:2143–2151
Sorenson C, Drummond M (2014) Improving medical device regulation: the United States and Europe in perspective. Milbank Q 92:114–150
Sowdhamini R, Srinivasan N, Shoichet B, Santi DV, Ramakrishnan C, Balaram P (1989) Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. Protein Eng 3:95–103
Sreedhara A, Yin J, Joyce M, Lau K, Wecksler AT, Deperalta G, Yi L, Wang YJ, Kabakoff B, Kishore RS (2016) Effect of ambient light on IgG1 monoclonal antibodies during drug product processing and development. Eur J Pharm Biopharm 100:38–46
Stefani M (2004) Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. BBA Mol Basis Dis 1739:5–25
Stone D (2010) Novel viral vector systems for gene therapy (molecular diversity preservation international). Viruses 2(4):1002–1007
Tamarappoo B, Yang, Verkman A (1999) Misfolding of mutant aquaporin-2 water channels in nephrogenic diabetes insipidus. J Biol Chem 274:34825–34831
Tandon R, Luxami V, Dosanjh HS, Tandon N, Paul K (2018) Insulin therapy for diabetes epidemic: a patent review. Curr Drug Deliv 15:777–794
Tatzelt J, Prusiner SB, Welch WJ (1996) Chemical chaperones interfere with the formation of scrapie prion protein. EMBO J 15:6363–6373
Telikepalli SN, Kumru OS, Kalonia C, Esfandiary R, Joshi SB, Middaugh CR, Volkin DB (2014) Structural characterization of IgG1 mAb aggregates and particles generated under various stress conditions. J Pharm Sci 103:796–809
Tokhadze N, Chennell P, Le Basle Y, Sautou V (2018) Stability of infliximab solutions in different temperature and dilution conditions. J Pharm Biomed Anal 150:386–395
Twomey A, Less R, Kurata K, Takamatsu H, Aksan A (2013) In situ spectroscopic quantification of protein–ice interactions. J Phys Chem B 117:7889–7897
Vestergaard B, Groenning M, Roessle M, Kastrup JS, Van De Weert M, Flink JM, Frokjaer S, Gajhede M, Svergun DI (2007) A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. PLoS Biol 5(5):e134
Vlieland ND, Nejadnik MR, Gardarsdottir H, Romeijn S, Sediq AS, Bouvy M, Egberts A, van den Bemt B, Jiskoot W (2018) The impact of inadequate temperature storage conditions on aggregate and particle formation in drugs containing tumor necrosis factor-alpha inhibitors. Pharm Res 35(2):42
Wang W (1999) Instability, stabilization, and formulation of liquid protein pharmaceuticals. Int J Pharm 185:129–188
Wang AY, Peng PD, Ehrhardt A, Storm TA, Kay MA (2004) Comparison of adenoviral and adeno-associated viral vectors for pancreatic gene delivery. Human Gene Therapy 15(4):405–413
Wang Z, Kuhr CS, Allen JM, Blankinship M, Gregorevic P, Chamberlain JS, Tapscott SJ, Storb R (2007) Sustained AAV-mediated dystrophin expression in a canine model of duchenne muscular dystrophy with a brief course of immunosuppression. Mol Ther 15(6):1160–1166
Wang W, Nema S, Teagarden D (2010) Protein aggregation—pathways and influencing factors. Int J Pharm 390:89–99
Wang S, Wu G, Zhang X, Tian Z, Zhang N, Hu T, Dai W, Qian F (2017a) Stabilizing two IgG1 monoclonal antibodies by surfactants: balance between aggregation prevention and structure perturbation. Eur J Pharm Biopharm 114:263–277
Wang S, Zhang X, Wu G, Tian Z, Qian F (2017b) Optimization of high-concentration endostatin formulation: harmonization of excipients' contributions on colloidal and conformational stabilities. Int J Pharm 530:173–186

Weiss M, Steiner DF, Philipson LH (2014) Insulin biosynthesis, secretion, structure, and structure-activity relationships. In: Feingold KR, Anawalt B, Boyce A et al (eds) Endotext [internet]. MDText.com, Inc., South Dartmouth

Welch WJ, Brown CR (1996) Influence of molecular and chemical chaperones on protein folding. Cell Stress Chaperones 1:109

White JR (2014) A brief history of the development of diabetes medications. Diabetes Spectr 27:82–86

Wilcox G (2005) Insulin and insulin resistance. Clin Biochem Rev 26:19–39

Wlodarczyk SR, Custódio D, Pessoa A Jr, Monteiro G (2018) Influence and effect of osmolytes in biopharmaceutical formulations. Eur J Pharm Biopharm 131:92–98

Yaffe K, Blackwell T, Whitmer R, Krueger K, BarrettConnor E (2006) Glycosylated hemoglobin level and development of mild cognitive impairment or dementia in older women. J Nutr Health Aging 10:293

Yancey PH (2005) Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. J Exp Biol 208:2819–2830

Yancey PH, Siebenaller JF (2015) Co-evolution of proteins and solutions: protein adaptation versus cytoprotective micromolecules and their roles in marine organisms. J Exp Biol 218:1880–1896

Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN (1982) Living with water stress: evolution of osmolyte systems. Science 217:1214–1222

Yao ZP, Zeng ZH, Li HM, Zhang Y, Feng YM, Wang DC (1999) Structure of an insulin dimer in an orthorhombic crystal: the structure analysis of a human insulin mutant (9 Ser→ Glu). Acta Crystallogr D Biol Crystallogr 55:1524–1532

Zhai J, Lee TH, Small DH, Aguilar MI (2012) Characterization of early stage intermediates in the nucleation phase of Aβ aggregation. Biochemistry 51:1070–1078

Zhang L, Graziano K, Pham T, Logsdon CD, Simeone DM (2001) Adenovirus-mediated gene transfer of dominant-negative Smad4 blocks TGF-β signaling in pancreatic acinar cells. Am J Physiol Gastrointest Liver Physiol 280(6):G1247–G1253

Zhang L, Yu L, ZhangVanEnk J, Huang G, Zhang J (2017) Phase behavior of an Fc-fusion protein reveals generic patterns of ion-specific perturbation on protein-protein interactions. J Pharm Sci 106:3287–3292

Zhou C, Qi W, Lewis EN, Randolph TW, Carpenter JF (2016) Reduced subvisible particle formation in lyophilized intravenous immunoglobulin formulations containing polysorbate 20. J Pharm Sci 105:2302–2309

Structural and Functional Aspects of Muscarinic Receptors in Correlation with Anticholinergic Drugs

13

Pramod Kumar Singh, Rajendra Nath, Ram Naraian, and Manish Kumar Gupta

Abstract

The muscarinic acetylcholine receptors (mAChRs) are receptors that produce the GPCR complex in the membrane of specific neurons and other cells. It performs a key role at the end of the receptor stimulated by the neurotransmitter. Ach liberates from postganglionic neurons in a parasympathetic region of ANS. The mAChRs constitute a family of five interrelated GPCRs that come under the category of α branch of GPCRs' Class A. The five different subtypes of the mAChR family are designated as M1–M5. M1, M3 and M5 subtype receptors exhibit to pair through the Gq/11 family of G proteins, but the M2 and M4 subtype receptors particularly indicate through Gi/o family of G protein. The mAChRs play multifunctional peripheral and central roles in human physiology including regulation of muscle contraction, heartbeat, lung, secretion by gland and other functions of the CNS.

Keywords

Muscarinic receptors · Acetylcholine receptor · Anticholinergic drugs · Central nervous system · *Atropa belladonna*

The original version of this chapter was revised due to the error in the affiliation of the co-authors. A correction to this chapter is available at https://doi.org/10.1007/978-981-15-5530-5_18

P. K. Singh
Department of Biochemistry, Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India

R. Nath
Department of Pharmacology and Therapeutics, King George's Medical University, Lucknow, India

R. Naraian · M. K. Gupta (✉)
Department of Biotechnology, Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_13

13.1 Introduction

Acetylcholine receptor (AChR) is an intrinsic membranous type of protein, which reacts to the binding of the neurotransmitter acetylcholine (ACh) molecule. It is classified into muscarinic and nicotinic receptors based on their pharmacology and relative target to molecules (Verma et al. 2018). Muscarinic and nicotinic receptors are the main type of the cholinergic system. The cholinergic system, portion of the visceral or autonomic nervous system (ANS), plays a significant role in many functions such as circadian rhythmicity, digestion, addiction, control of heartbeat, motivation, blood pressure, cognitive flexibility, pain and reward, spatial learning and perceptual memory (Prado et al. 2017). mAChRs are well-known metabotropic acetylcholine receptors that are mainly reactive to muscarine. mAChRs are termed after muscarine, a lethal alkaloid produced by the highly poisonous mushroom *Amanita muscaria* (Jo et al. 2014). Scopolamine and atropine are the best known naturally occurring muscarinic antagonist, which is reported in the fatal nightshade plant: *Atropa belladonna* (Albuquerque et al. 2009). Nicotinic acetylcholine receptors (nAChRs) are famous ionotropic acetylcholine receptors particularly responsive to nicotine, Na^+, Ca^{2+} and K^+ ion channel (Corradi and Bouzat 2016). nAChR is named after nicotine, an ideal agonist. D-tubocurarine compound, a toxic alkaloid isolated from the curare poison, is a very well-known nicotinic antagonist (Malca Garcia et al. 2015) (Fig. 13.1).

13.1.1 Structure and Function of Muscarinic Receptor and Their Subtypes

The mAChRs are acetylcholine receptors, which produce the GPCR complex in the membrane of specific neurons and other cells (Eglen 2006). It performs a key role at the end of the receptor stimulated by the neurotransmitter. Ach liberates from postganglionic neurons in a parasympathetic region of ANS. The mAChRs constitute a family of five interrelated GPCRs that comes under the category of α branch of GPCRs Class A (Fredriksson et al. 2003). The five different subtypes of the mAChR family are designated as M1–M5 (encoded by CHRM1–CHRM5 genes). M1, M3 and M5 subtype receptors exhibit to pair through the Gq/11 family of G proteins, but the M2 and M4 subtype receptors particularly indicate through Gi/o family of G protein (Haga 2013). The mAChRs play multifunctional peripheral and central roles

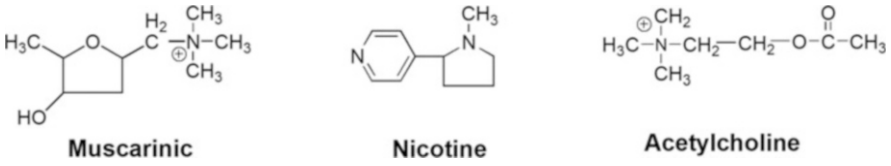


Fig. 13.1 Chemical structures of muscarine, acetylcholine and nicotine

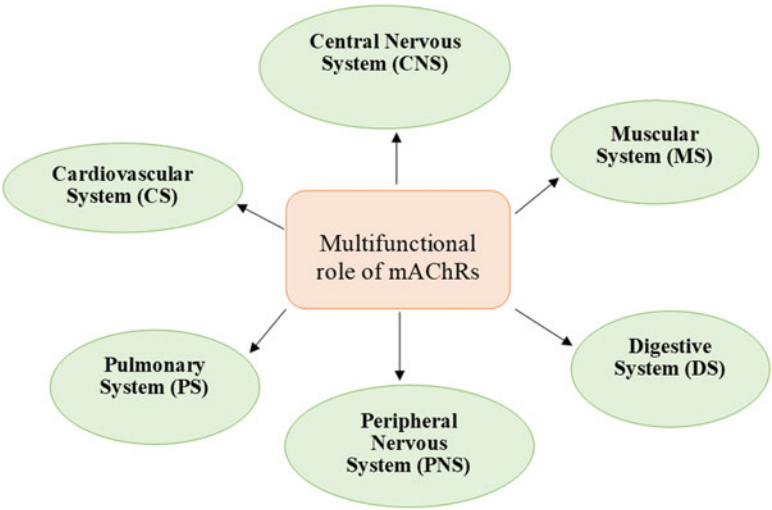


Fig. 13.2 Multifunctional role of mAChRs

in human physiology including regulation of muscle contraction, heartbeat, lung, secretion by gland and other functions of the CNS (Wess et al. 2007) (Fig. 13.2).

13.1.1.1 Muscarinic-1 Receptor (M1 Receptor)

Muscarinic-1 receptor (M1 receptor) is a cholinergic muscarinic type of receptor found in humans, rats and mice encoded by the CHRM1 gene (CHRM1 2020). The receptor belongs to the GPCR family and bound to Gq proteins (Qin et al. 2011). This is one of the five muscarinic receptors that act as the metabotropic roles of ACh in the CNS of humans. M1 receptors mostly found in nerve cells of the hippocampus and cerebral cortex. Initiation of the M1 receptor yields many reactions including the activation of ion channels such as Cl^-, K^+, inhibition of cAMP production and the upregulation of phospholipase C (Sanchez et al. 2009). M1 receptor agonists may also lead to secretion from the bronchoconstriction, stomach and salivary gland. M1 receptors generally participate in many processes including cardiac muscle contraction, control of seizure and cognitive activity (Hamilton et al. 2001; Bakker et al. 2018). The beginning of these receptors by selective agonists reduces harmful β-amyloid secretion and increases the secretion of the non-toxic α-amyloid peptide from amyloid precursor proteins (Jiang et al. 2014).

The M1 receptor is made up of 521 amino acid residues (Fig. 13.3). These are made up of five transmembrane domains: residues 1–239, 403–515, 240–255, 298–402, 256–298, respectively. The M1 muscarinic receptor interacts with the inhibitor tiotropium. Orthosteric and allosteric interaction sites play a significant role in drug specificity. It also reveals how allosteric modulation may be spread involving the two spatially discrete domains (Thal et al. 2016).

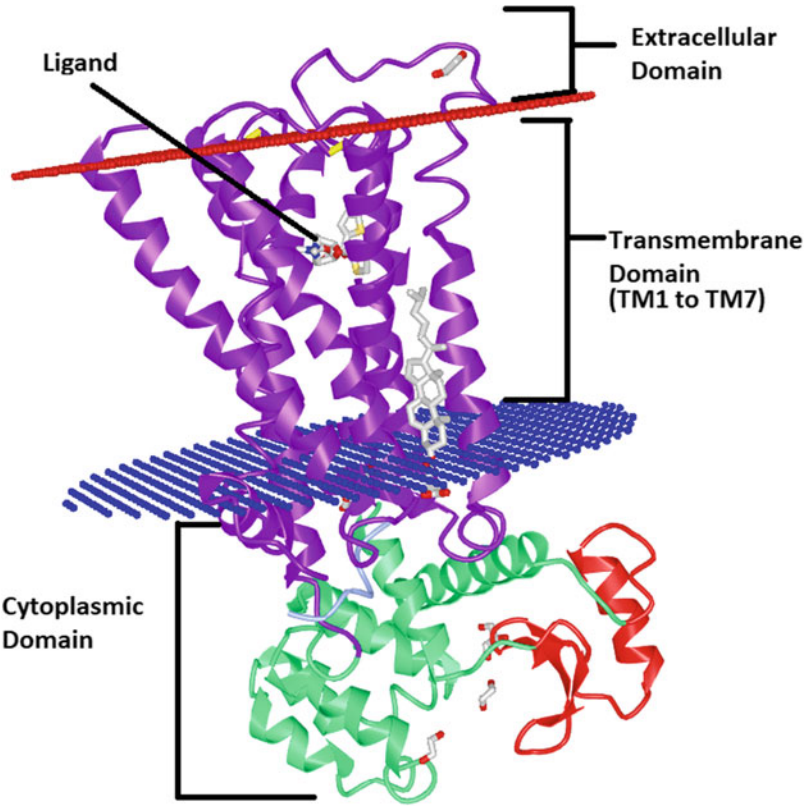


Fig. 13.3 Structure of muscarinic receptor-1 (PDB Id: 5CXV)

13.1.1.2 Muscarinic-2 Receptor (M2 Receptor)

The mAChR M2 is also known as the cholinergic receptor M2, which is encoded by the CHRM2 gene of mice, rats and humans (CHRM2 2020). M2 receptor is a member of the GPCRs family that binds to Gi protein, generally leading to inhibitory effects (Douglas et al. 2001). It regulates the metabotropic function of ACh in the CNS. The receptor is tightly engaged in brain regions, heart and smooth muscle. Initiation of the M2 receptor produces several responses such as initiation of Ca^{2+}, K^+ channels and the inhibition of adenylyl cyclase (Harvey and Belevych 2003). M2 receptor also participates in several processes such as regulation of atrial contraction, AV node conduction velocity, acquiring and retention of smooth muscle contraction (Andersson and Olshansky 2007). The receptor antagonist has been suggested useful in the remedy of Alzheimer's disease (Clader and Wang 2005; Kumar et al. 2016; Wang et al. 2020) (Fig. 13.4).

The receptor is made up of 467 amino acids with 68% helical (24 helices; 322 residues) and 2% beta sheet (4 strands; 13 residues). M2 receptor is devoid of the third intracellular loop and the natural glycosylation sites in the majority of cases.

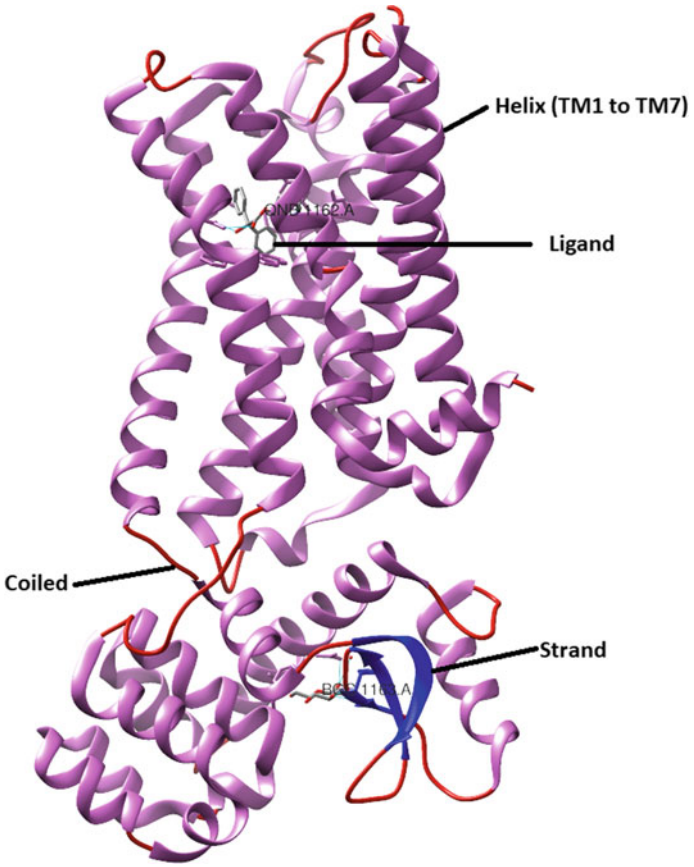


Fig. 13.4 Structure of muscarinic receptor-2 (PDB ID: 3UON)

There are ample hydrophobic interactions that occur between receptor proteins within the transmembrane. The ligand QNB in the interior buried pocket specified by the side chains of TM3–TM7 (Haga et al. 2012). A hydrophobic layer formed by three amino acids, viz. Leu 65 in TM2, Leu 114 in TM4 and Ile 392 in TM6. The orthosteric binding pocket is produced by residues that are identical in M1–M5 receptors. All the muscarinic receptors (M1–M5) show common structural homology with other activity distinct acetylcholine interacting proteins from diverse species. M2 receptor structure imparts molecular insights into the contests of creating specific ligands for muscarinic receptors and their predisposition for allosteric control.

13.1.1.3 Muscarinic-3 Receptor (M3 Receptor)

The mAChR M3 is known as acetylcholine/cholinergic receptor M3, which is encoded by the CHRM3 gene of the mouse, rats and humans (CHRM3 2020). M3 receptor is a member of the GPCRs family that binds to Gq protein, which

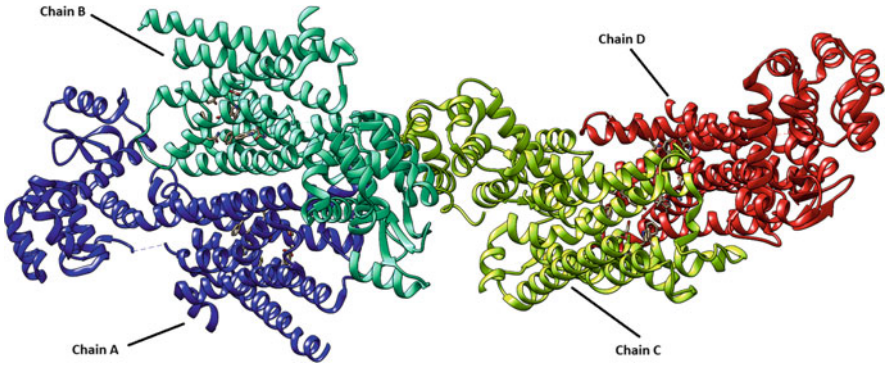


Fig. 13.5 Structure of muscarinic receptor-3 (PDB Id: 4DAJ)

upregulates inositol triphosphate (IP3) and phospholipase C and increases the intracellular Ca^{2+} (Qin et al. 2011). It acts as the metabotropic role of ACh in the CNS. These receptors are generally found in the lungs, endocrine, exocrine glands, smooth muscle and CNS (Weston et al. 2012). The receptor agonist performs an important function in bronchoconstriction and smooth muscle constriction. Activation of the M3 receptor leads to various secretions from the pancreas, stomach and salivary gland (Gautam et al. 2006). Therefore, the M3 receptor actively participates in many metabolic activities such as regulation of the insulin release and the glucose homeostasis. Additionally, the receptor is a potentially beneficial target site in the case of the pulmonary block and in the progression of colon cancer (Moulton and Fryer 2011; Tolaymat et al. 2019). Moreover, initiation of the M3 receptor by selective agonists may also be useful in the case of type-2 diabetes (Gautam et al. 2006; Ito et al. 2019).

The M3 receptor is made up of 479 amino acid residues having four chains A, B, C, D and two domains (Fig. 13.5). The domain 1 lies from 1–202 to 368–479 range, whereas domain 2 ranges from 203 to 367 amino acid residues. No structure alignment results are available for all the four chains of PDB ID: 4DAJ (A to D) explicitly. Structural conservation comprises of intracellular loops 1 and 2. It also includes extracellular loops 1, 2 and 3, having extremely common resemblance in character; overall folds even though little sequence conservation. The M3 receptor shows distinctive characteristics, having a big extracellular vestibule as part of a lengthened hydrophilic route comprising the orthosteric binding pocket (Kruse et al. 2012). Molecular dynamics simulations study advocates that ligand tiotropium interacts momentarily to an allosteric site on the way to the cavity in the interior of the receptor that possesses suitable properties for binding a ligand. The binding pocket of the receptor may also provide an opportunity to design a novel ligand with enhanced therapeutics for the M3 receptors. The conserved residue (Thr234 of TM 5 and Tyr506 of TM 6) in all muscarinic receptors may perform a significant role in the designing of a novel stimulator for the activation of muscarinic receptors (Wess et al. 1992).

13.1.1.4 Muscarinic 4 Receptor (M4 Receptor)

The mAChR M4 receptor is also known as the cholinergic receptor M4. M4 receptor is present in rats, humans and mice, which is encoded by the CHRM4 gene (CHRM4 2020; Birdsall et al. 2019). This receptor is a participant of the GPCRs family, mostly bind to Gi proteins, leading to inhibitory effects (Douglas et al. 2001). It regulates the metabotropic functions of acetylcholine in the brain. This receptor is tightly involved in the lung and striatum. Activation of the M4 receptor response to several reactions includes the inhibition of adenylyl cyclase (Guo et al. 2010). The function of the M4 receptor is the indirect mediation of dopaminergic neurotransmission through cholinergic activity. M4 receptor is also reported to be involved in neuropathological diseases (Tzavara et al. 2004; Stepnicki et al. 2018).

The M4 receptor is made up of two domains and two chains (A & B). The first domain ranges from residues 1–204 to 326–422 and domain 2 ranges from residues 205 to 325 (Fig. 13.6). An alteration in the rotamer of D112 amino acid transmembrane 3 is conserved all over the biogenic amine G and acts as the counter ion for positively charged neurotransmitters (Van Rhee and Jacobson 1996). This rotameric alteration indicates that D112 of TM3 is beyond ligand tiotropium. The residues Y439 and Y443 play a significant affair to stabilize the various inoperative states of conformation to ligand interact with it. The orthosteric site of the M4 receptor is nearer to the M1 than the M2 subtypes. A deviation in amino acids covering allosteric site stresses the significance of this zone for designing specific drugs (Thal et al. 2016).

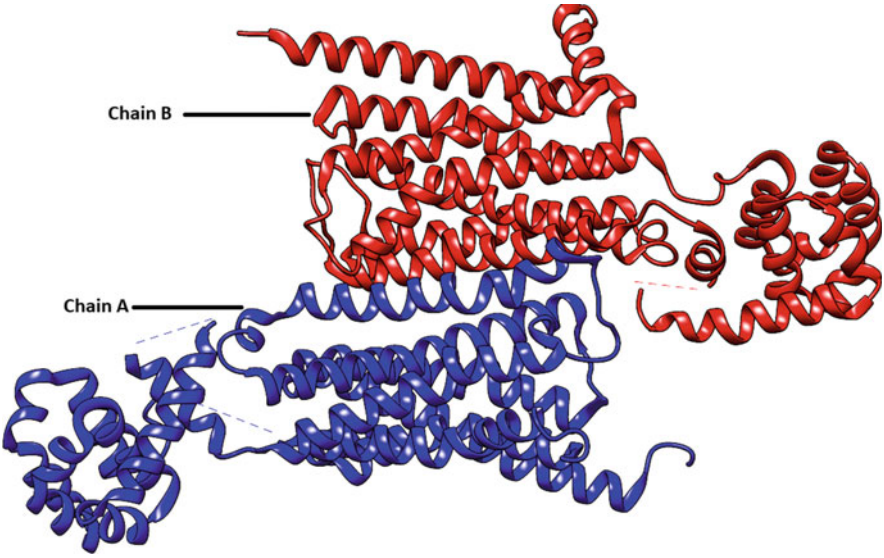


Fig. 13.6 Structure of muscarinic receptor-4 (PDB Id: 5DSG)

13.1.1.5 Muscarinic 5 Receptor (M5 Receptor)

The M5 receptor is encoded by the CHRM5 gene of rats, mouse and human, which is a member of the GPCRs subfamily of the integral membrane protein (CHRM5 2020). The receptor is coupled to Gq protein (Qin et al. 2011). It regulates the metabotropic function of ACh in the CNS. M5 receptor is the most closely occupied in the neurotransmitter containing neuronal cells in the cerebral cortex, striatum, hippocampus along substantia nigra of the brain (Foster et al. 2014). M5 receptor agonists regulate the level of the dopaminergic neuron and release dopamine into the striatum, which facilitates rude substances such as cocaine (Fink-Jensen et al. 2003). The clinical effect of this receptor is not very well-known; however, activation of the M5 receptor is identified which reduces the level of cyclic AMP and the activities of protein kinase C (Bender et al. 2018). M5 receptors may also probably beneficial in the treatment of memory deficits produced by diminished cerebrovascular function (Araya et al. 2006; Vuckovic et al. 2019).

The experimental structure of the M5 receptor is not available till date. M5 receptor participates in various cellular activities such as K^+ channel modulation, phosphoinositide degradation and adenylate cyclase inhibition (UniprotID: P08912). All five subtypes (M1–M5) of mAChRs play a significant role in biological processes like renal, cardiac, intestinal function, motor control, cognitive and attention mechanisms. Different functions and diseases associated with these five subtypes along with their locations are given in Table 13.1.

13.2 Anticholinergic Drugs

Anticholinergic drugs are substances which inhibit the action of acetylcholine at the synapse of the PNS and CNS (Xu et al. 2017). It blocks the parasympathetic nerve impulse by non-selectively or selectively linking the neurotransmitter acetylcholine to its receptor site of neurons (Prommer 2013). These drugs are also called parasympatholytics or cholinergic antagonists. Anticholinergics drugs are divided into two main categories based on their specific target to PNS and CNS: antinicotinic drugs and antimuscarinic drugs.

13.2.1 Antinicotinic Drugs

The antinicotinic drugs attack on the nAChRs. The majority of antinicotinic drugs are non-depolarizing and depolarizing drivers. Non-depolarizing agents such as vecuronium, rocuronium, pancuronium, cisatracurium, atracurium and mivacurium are a type of neuromuscular blocker, which does not depolarize the motor end plate and causing action potential (Kim et al. 2017). Depolarizing agents such as succinylcholine, decamethonium, and others are a type of neuromuscular blocker which depolarize the motor end plate and produce an action potential (Ahmad et al. 2018). Both agents are used in muscle relaxants for clinical purposes (Clar and Liu 2020; Gulenay and Mathai 2020).

Table 13.1 Location and function of M1–M5 receptor and their associated diseases

S. no.	mAChRs subtypes	Brain regions	Other tissue	Function	Associated diseases	References
1.	M1	Rich in hippocampus, neostriatum, cerebral cortex, and constitute 50–60% of total acetylcholine receptor	Sympathetic ganglia, salivary gland	Cognition, salivation	Alzheimer's disease, cognitive dysfunction	Lee et al. (2016); Lebois et al. (2018)
2.	M2	Situated throughout brain	Cardiac muscle, smooth muscle	Cardiac inhibition	Cardiac dysfunction, pain	Saternos et al. (2018); Kobayashi et al. (2007)
3.	M3	Low levels throughout brain	Salivary glands, smooth muscle, eyes	Contraction of blood vessel, lachrymal secretion, regulation of GIT	COPD, irritable bowel syndrome, urinary incontinence	Saternos et al. (2018); Kobayashi et al. (2007); Muise et al. (2017)
4.	M4	Rich in hippocampus, cortex and neostriatum	Salivary gland	Regulatory action on Ca^{+2} and K^{+} channel	Schizophrenia, Parkinson's disease, neuropathic pain	Lebois et al. (2018); Langmead et al. (2008); Kobayashi et al. (2007)
5.	M5	Estimate neuron of hippocampus, substantia nigra, pars compacta and ventral tegmental region	Muscle (eye)	Regulation of dopamine liberation at terminal end inside substantia nigra	Parkinson's disease, schizophrenia, drug addiction	Lebois et al. (2018); Langmead et al. (2008)

13.2.2 Antimuscarinic Drugs

Antimuscarinic drugs act on the mAChRs. Muscarinic antagonists (muscarinic anticholinergic drugs) disrupt the learning and memory processes (Table 13.2). These drugs are involved in causing cognitive and memory deficits in an experimental animal model for the pathological conditions identified in several human neuropathological diseases such as Alzheimer's, Schizophrenia and other diseases (Robinson et al. 2011). Majority of anticholinergic drugs have been used in a wide range of clinical conditions like amnesia, mydriasis, bronchodilation and sedation (Prommer 2013).

13.3 Sources of Antimuscarinic Drugs

The most common sources of anticholinergic drugs are (1) *Datura species* (Datura 2016), (2) *Atropabelladonna* (Ulbricht et al. 2004; Belladonna 2020), (3) *Hyoscyamusniger* (Roberts and Wink 1999), (4) *Brugmansia species* (toxic plants 2020), (5) *Garrya species* (Nesom 2012) (6) and *Mandragora officinarum* (Duke 2002) plants.

13.4 Classification of Antimuscarinic Drugs

13.4.1 Based on Their Sources

Antimuscarinic drugs are classified into three groups (Fig. 13.7): (1) natural alkaloids, (2) semi-synthetic drugs and (3) synthetic drugs.

13.4.1.1 Natural Alkaloids

Natural alkaloids are mostly natural organic compounds, which usually consist of a basic nitrogen atom. It also includes some correlated compounds with both neutral and weak acidic in nature (IUPAC 2012; Sheela 2013). Some alkaloids are synthetic or semisynthetic compounds of dissimilar or similar structures like natural alkaloids. Alkaloids have various important physiological roles in humans and animals (Lahlou 2014). Natural alkaloids such as scopolamine, atropine and tubocurarine are well known and may be toxic to the animal. They exhibit a broad range of pharmacological and biological properties (Yadav et al. 2014).

13.4.1.2 Semi-Synthetic Drugs

Semi-synthetic drugs are produced by the biochemical reaction between naturally occurring compounds to form a new product (Lahlou 2014). Semi-synthesis, a kind of biochemical synthesis, which uses chemical compound extracted from a natural source (plant material or microbial cell cultures) as the initial materials to yield other innovative compounds (Cragg and Newman 2013). These drugs are neither synthetic nor natural completely, which is a mixture of both. Semi-synthetic drugs are

Table 13.2 Features of muscarinic receptor antagonists and their functions

S. no.	Characteristics	Antagonists	Functions	References
1.	Tertiary amines	Atropine	Increases heart rate.	McEvoy (2018); De Caen et al. (2015)
	Lipophilic (good oral bioavailability and CNS penetration)		Diminishes secretions of exocrine glands.	
			Diminishes motility and tone of smooth muscle.	
			Diminishes cholinergic overactivity in the brain.	
			Mydriasis and Cycloplegia	
		Scopolamine (hyoscine)	CNS depression	McEvoy (2005); Rang (2003)
			Diminishes vestibular disturbances (antiemetic)	
		Homatropine	Mydriasis	Agrawal et al. (2010); Yazdani et al. (2018)
		Tropicamide	Impair accommodation	
		Cyclopentolate		
		Benztropine	Diminishes cholinergic overactivity in CNS	McEvoy (2003); Harvey et al. (2018)
		Biperiden		
		Trihexyphenidyl		
		Tolterodine	Diminishes motility and tone of muscle cells	De Caen et al. (2015); Katzung and Trevor (2014)
		Oxybutynin		
		Solifenacin		
		Dicyclomine		
		Darifenacin	Enhances sphincter tone	De Maagd and Davenport (2012)
2.	Quaternary amines hydrophilic (less oral bioavailability and CNS penetration)	Butyl scopolamine (hyoscine butyl bromide)	Reduces motility and tone of the gut (antispasmodic effects)	Tytgat (2007)
		Methscopolamine	Diminishes secretion of exocrine gland	Ivanovic et al. (2016)
		Glycopyrrolate		
		Pirenzepine		
		Propantheline		
		Ipratropium bromide	Bronchodilation	Rang (2003)
		Tiotropium bromide		Cheyne et al. (2013)

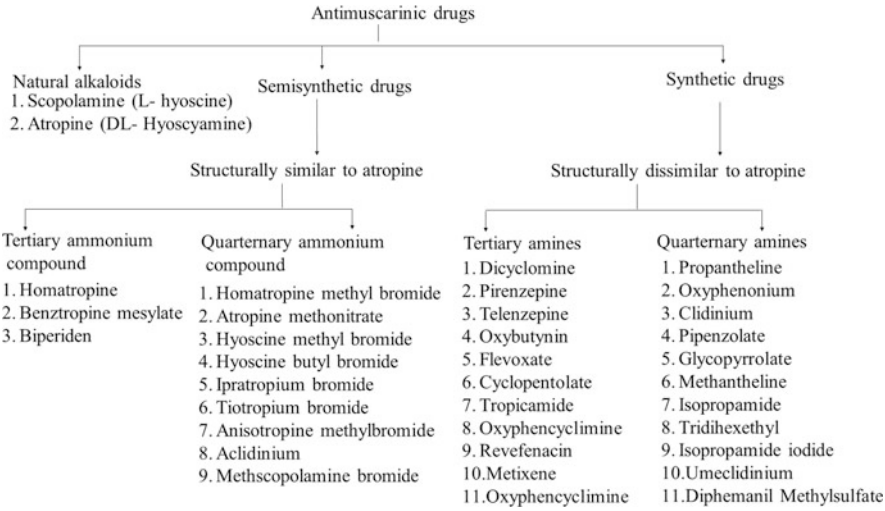


Fig. 13.7 Classification of antimuscarinic drugs: natural, semisynthetic and synthetic muscarinic receptor antagonist

structurally related to atropine including tertiary ammonium compound (homatropine, biperiden, etc.), quaternary ammonium compound (homatropine methyl bromide, atropine methonitrate, etc.) and novel antibiotics (tetracycline, doxycycline, tigecycline and chemotherapy drug) which exhibit a wide variety of chemical and pharmacological properties (Lahlou 2014; Nelson and Levy 2011; Liu and Myers 2016).

13.4.1.3 Synthetic Drugs

The synthetic drug is a drug having similar properties and belongings to hallucinogen or narcotic drugs (Garcia-Romeu et al. 2016). These drugs are structurally dissimilar to atropine including tertiary amines (pirenzepine, dicyclomine, oxybutynin, etc.) and quaternary amine (propantheline, glycopyrrolate, trihexyphenidyl, etc.) which exhibits a wide variety of biological, physicochemical and pharmacological properties (Grynkiewicz and Gadzikowska 2008). Synthetic drugs are prepared from the beginning substance, which is not present in the environment; instead, they are formed from building blocks of a chemical substance (Lahlou 2014). The common process of synthetic drug discovery is analogous to the discovery of natural drugs (Mathur and Hoskins 2017). In the natural drugs discovery, the compounds are obtained from sources like the plant, animal and microorganism. In the case of synthetic drug discovery, the compound generally produces in vitro laboratory through combinatorial technique, which manufactured a hundred to million molecules from the building block of smaller chemical substances (Valecha et al. 2010).

13.4.2 Based on Their Mode of Action

Antimuscarinic drugs are classified into two groups (Fig. 13.8): (1) nonselective muscarinic receptor antagonist and (2) selective muscarinic receptor antagonist.

13.4.2.1 Nonselective Muscarinic Receptor Antagonist

The nonselective muscarinic antagonist is a drug, which is not selective for all subtypes of the muscarinic receptors on therapeutic doses (Svoboda et al. 2017). Most of the anticholinergic drugs such as scopolamine, atropine and homatropine are nonselective for the subtypes M1–M5 receptors. However, these drugs are specific to the muscarinic receptor (Svoboda et al. 2017). A nonselective muscarinic drug is used as a medication for clinical events like obstruction of muscle contraction, salivary secretion and cardio-protection; conversely, their beneficial function in the treatment of long standing is known (Chapple et al. 2002). Ipratropium and oxitropium are also nonselective antimuscarinic drugs that successfully retract airway hyperactivity and bronchoconstriction in humans (Coulson and Fryer 2003).

13.4.2.2 Selective Muscarinic Receptor Antagonists

The muscarinic receptor is well-defined by selective agonists and antagonists. A selective muscarinic receptor grouping preceded the identification of acetylcholine (Schiechl et al. 2008). Telenzepine and pirenzepine antagonists have a comparatively strong binding affinity for the M1 receptor, which permitted for use in the therapy of peptic ulcer disease (Okabe et al. 2002). Conversely, M2 receptor antagonist includes Otenzepad peripherally acting in the remedy of bradycardia (Lanzafame et al. 2001). Darifenacin inhibitor is applied in the remedy of irritable bowel syndrome, urinary incontinence and is a specific M3 receptor (McFerren and Gomelsky 2015). Biperiden, a comparative specific M1 receptor antagonist, therapeutically applied to reduce the symptoms of Parkinson's disease, memory and learning deficit in Alzheimer's disease (Witkin et al. 2014).

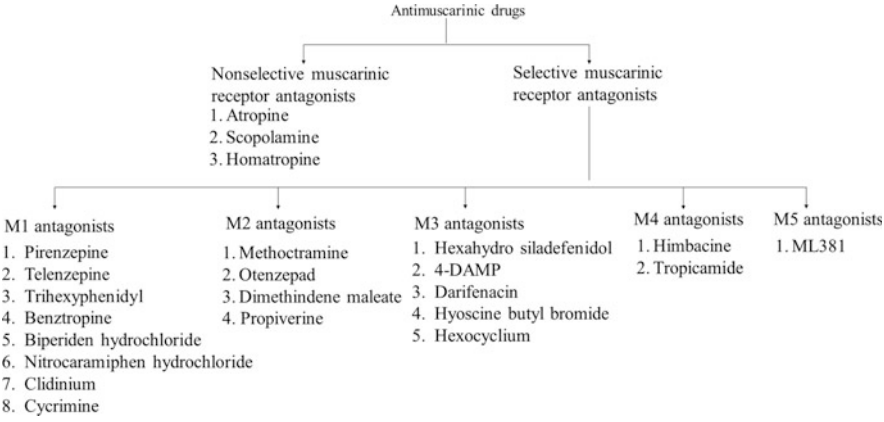


Fig. 13.8 Classification of antimuscarinic drugs: nonselective and selective receptor antagonist

A recent study was done in the chick model of myopia, which proved that the best selective M4 receptor antagonist himbacine is the most effective drug in the regulation of myopia (Carr et al. 2019; Cottriall et al. 2001). However, the M5 receptor has least worked out due to the absence of selective ligand subtypes of mAChRs. After the detection, characterization and synthesis of the first highly muscarinic and specific M5 orthosteric inhibitor, ML381 or VU0480131 have been reported (Gentry et al. 2014).

13.5 Mechanism of Antimuscarinic Drugs

The antimuscarinic drugs including atropine, scopolamine and others are more liposoluble because of their lipophilic nature, which act rapidly and are being absorbed from the gastrointestinal tract (GIT). However, it is less absorbed from injured or intact skin and easier to cross the blood–brain barrier (BBB), which upset the CNS and other organ systems (Rajput 2013; He et al. 2011). Most of the antimuscarinic drugs are usually observed to be safe taking at dose level 1.5 mg/day (Beyer et al. 2009; Ulbricht et al. 2004). Toxicity usually occurs after the ingestion of drugs at a dose level more than 1.5 mg/day in the brain, which caused unclear vision, delirium, incomprehensive speech, fatigue and unconsciousness (Milanlioglu 2011; Apfel et al. 2010; Bogan et al. 2009).

Antimuscarinic drugs competitively bind and inhibit acetylcholine from the binding site of the muscarinic receptors (Pergolizzi et al. 2012). However, their antagonistic actions may be decreased by elevating the concentration of the muscarinic agonists. The main action of antimuscarinic drugs like scopolamine, atropine and associated drugs competitively blocks the action of ACh agonists (Fig. 13.9). These drugs compete with such agonists for normal requisite on the muscarinic receptor. Many evidences support the idea that scopolamine- and atropine-related compounds compete with agonists for normal requisite on the muscarinic receptors (Snyder et al. 2005; Malik et al. 2015).

13.5.1 Epidemiology of Anticholinergic Drug

The epidemiological study suggests that about 20–50% of individuals of old age are regularly put in danger to anticholinergic drugs with possible activity (Fox et al. 2011). This shows that more than one half of the drug usually given for grown-up people is possible due to anticholinergic action (Chew et al. 2008). Anticholinergic agent's actions in adult individuals differ with sex, age and comorbidities (Wawruch et al. 2012; Chatterjee et al. 2010; Agar et al. 2009).

13.5.2 Clinical Significance of Anticholinergic Drugs

Anticholinergic drugs with potential properties have been significantly used in medicine for many years to treat disease conditions including the following:

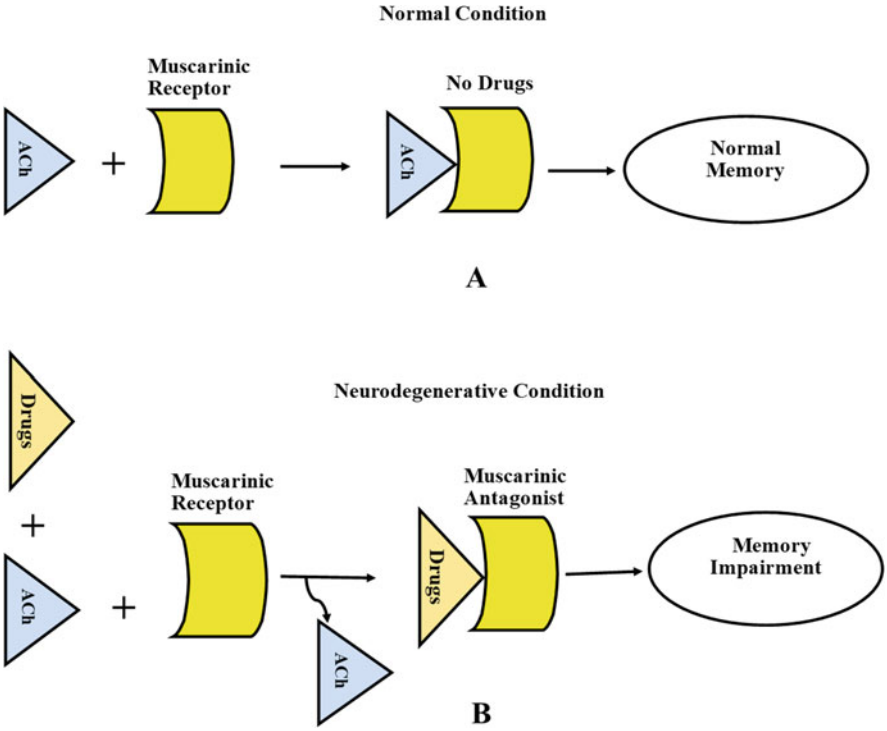


Fig. 13.9 Mechanism of action of agonist and antimuscarinic drugs on muscarinic receptor site: (**a**) normal condition and (**b**) neurodegenerative condition

(1) motion sickness, (2) Parkinson's disease, (3) overactive bladder and urinary incontinence, (4) psychiatric disorders, (5) gastrointestinal disorders, (6) diarrhoea, (7) asthma, (8) chronic obstructive pulmonary disease (COPD), (9) surgery and anaesthesia for muscle relaxation, (10) anaesthesia during surgery, (11) insomnia, (12) Alzheimer's disease and (13) toxicity of certain poisonings (Cahalan et al. 2009; Kees et al. 2015).

13.5.3 Side Effects of Anticholinergic Drugs

Different studies have indicated the side effects of anticholinergic drugs on different organ systems of the human body. The highly significant side effects of anticholinergic drugs are mentioned in Table 13.3.

Table 13.3 Side effects of anticholinergic drugs on organs and other tissues

S. no.	Organ/tissue	Side effects	References
1.	Exocrine glands	Reduce respiratory tract secretions	Le and Bhushan (2014); Macdiarmid (2008)
		Sore throat and dry mouth	
		Dry skin, warm and hyperthermia	
2.	Cardiovascular system	Elevate heart rate (tachycardia)	Katzung and Trevor (2014); Cetinel and Onal (2013)
3.	Smooth muscle	Obstipation or ileus	Lieberman (2004); Mintzer and Burns (2000)
		Gastroesophageal reflux	
		Urinary retention/impaired maturation	
		Flush and vasodilatation	
4.	Eye	Photophobia and mydriasis	Mintzer and Burns (2000); Ramnarine (2020)
		Blurred vision	
5.	CNS	Agitation, hallucination and excitement together with lipophilic antimuscarinic drugs such as atropine and scopolamine, disorientation, confusion, seizure, coma and rarely death (elderly patients)	Cetinel and Onal (2013); Berdai et al. (2012)

13.6 Conclusions

The muscarinic receptor can be distinguished structurally, physiologically and pharmacologically. The muscarinic receptor is a family of GPCRs and commonly distributed in the human body. Each subtype performs a specific function and plays a significant biological activity in the PNS and CNS. Muscarinic receptor antagonists competitively inhibit postganglionic muscarinic receptor which regulates several essential functions and structures of the PNS and CNS. Thus, antagonistic activity of antimuscarinic drug acts as a target in correlation with muscarinic acetylcholine receptor subtypes in animal and is associated with human health risk.

References

Agar M, Currow D, Plummer J, Seidel R, Carnahan R, Abernethy AP (2009) Changes in anticholinergic load from regular prescribed medications in palliative care as death approaches. Palliat Med 36(3):257–265

Agrawal RV, Murthy S, Sangwan V, Biswas J (2010) Current approach in diagnosis and management of anterior uveitis. Indian J Ophthalmol 58(1):11–19

Ahmad M, Khan NA, Furqan A (2018) Comparing the functional outcome of different dose regimes of succinylcholine. When used for rapid induction and intubation. J Ayub Med Coll Abbottabad 30(3):401–404

Albuquerque EX, Pereira EFR, Alkondon M, Rogers SW (2009) Mammalian nicotinic acetylcholine receptors: from structure to function. Physiol Rev 89:73–120

Andersson KE, Olshansky B (2007) Treating patients with overactive bladder syndrome with antimuscarinics: heart rate considerations. BJU Int 100(5):1007–1014

Apfel CC, Zhang K, George E, Shi S, Jalota L, Hornuss C, Fero KE, Heidrich F, Pergolizzi JV, Cakmakkaya OS, Kranke P (2010) Transdermal scopolamine for the prevention of postoperative nausea and vomiting: asystematic review and meta-analysis. Clin Ther 32:1987–2002

Araya R, Noguchi T, Yuhki M, Kitamura N, Higuchi M, Saido TC, Seki K, Itohara S, Kawano M, Tanemura K, Takashima A, Yamada K, Kondoh Y, Kanno I, Wess J, Yamada M (2006) Loss of M5 muscarinic acetylcholine receptors leads to cerebrovascular and neuronal abnormalities and cognitive deficits in mice. Neurobiol Dis 24(2):334–344

Bakker G, Vingerhoets C, Boucherie D, Caan M, Bloemen O, Eersels J, Booij J, van Amelsvoort T (2018) Relationship between muscarinic M1 receptor binding and cognition in medication-free subjects with psychosis. Neuroimage Clin 18:713–719

Belladonna (2020) MedlinePlus, US national library of medicine. https://medlineplus.gov/druginfo/natural/531.html. Accessed Mar 2020

Bender AM, Cho HP, Nance KD, Lingenfelter KS, Luscombe VB, Gentry PR, Voigtritter K, Berizzi AE, Sexton PM, Langmead CJ, Christopoulos A, Locuson CW, Bridges TM, Chang S, O'Neill JC, Zhan X, Niswender CM, Jones CK, Conn PJ, Lindsley CW (2018) Discovery and optimization of potent and CNS penetrant M5-preferring positive allosteric modulators derived from a novel, chiral *N*-(Indanyl) piperidine amide scaffold. ACS Chem Neurosci 9 (7):1572–1581

Berdai MA, Labib S, Chetouani K, Harandou M (2012) Atropa belladonna intoxication: a case report. Pan Afr Med J 11:72

Beyer J, Drummer OH, Maurer HH (2009) Analysis of toxic alkaloids in body samples. Forensic Sci Int 185:1–9

Birdsall NJM, Bradley S, Brown DA, Buckley NJ, Challiss RJ, Christopoulos A, Eglen RM, Ehlert F, Felder CC, Hammer R, Kilbinger HJ, Lambrecht G, Langmead C, Mitchelson F, Mutschler E, Nathanson NM, Schwarz RD, Tobin AB, Valant C, Wess J. (2019) Acetylcholine receptors (muscarinic) (version 2019.4) in the IUPHAR/BPS Guide to pharmacology database. IUPHAR/BPS Guide to pharmacology CITE, 2019(4). http://journals.ed.ac.uk/gtopdb-cite/article/view/3156

Bogan R, Zimmermann T, Zilker T, Eyer F, Thiermann H (2009) Plasma level of atropine after accidental ingestion of Atropa belladonna. Clin Toxicol 47:602–604

Cahalan MD, Barash PG, Cullen BF, Stoelting RK (2009) Clinical anesthesia. Lippincott Williams & Wilkins, Hagerstwon MD

Carr BJ, Nguyen CT, Stell WK (2019) Alpha2-adrenoceptor agonists inhibit form-deprivation myopia in the chick. Clin Exp Optom 102(4):418–425

Cetinel B, Onal B (2013) Rationale for the use of anticholinergic agents in overactive bladder with regard to central nervous system and cardiovascular system side effects. Korean J Urol 54 (12):806–815

Chapple CR, Yamanishi T, Chess-Williams R (2002) Muscarinic receptor subtypes and management of the overactive bladder. Urology 60(5):82–88

Chatterjee S, Mehta S, Sherer JT, Aparasu RR (2010) Prevalence and predictors of anticholinergic medication use in elderly nursing home residents with dementia: analysis of data from the 2004 national nursing home survey. Drugs Aging 27(12):987–997

Chew ML, Mulsant BH, Pollock BG, Lehman ME, Greenspan A, Mahmoud RA, Kirshner MA, Sorisio DA, Bies RR, Gharabawi G (2008) Anticholinergic activity of 107 medications commonly used by older adults. J Am Geriatr Soc 56(7):1333–1341

Cheyne L, Irvin-Sellers MJ, White J (2013) Tiotropium versus ipratropium bromide for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 9(9):CD009552

CHRM1 (2020) Cholinergic receptor muscarinic 1 (Gene ID: 1128). https://www.ncbi.nlm.nih.gov/gene/1128. Accessed Mar 2020

CHRM2 (2020) Cholinergic receptor muscarinic 2 (Gene ID: 1129). https://www.ncbi.nlm.nih.gov/gene/1129. Accessed Mar 2020

CHRM3 (2020) Cholinergic receptor muscarinic 3 (Gene ID: 1131). https://www.ncbi.nlm.nih.gov/gene/1131. Accessed Mar 2020
CHRM4 (2020) Cholinergic receptor muscarinic 4 (Gene ID: 1132). https://www.ncbi.nlm.nih.gov/gene/1132. Accessed Mar 2019
CHRM5 (2020) Cholinergic receptor muscarinic 5 (Gene ID: 1133). https://www.ncbi.nlm.nih.gov/gene/1133. Accessed Mar 2020
Clader JW, Wang Y (2005) Muscarinic receptor agonists and antagonists in the treatment of Alzheimer's disease. Curr Pharm Des 11(26):3353–3361
Clar DT, Liu M (2020) Non-depolarizing neuromuscular blockers. Stat Pearls Publishing, Treasure Island
Corradi J, Bouzat C (2016) Understanding the bases of function and modulation of a7 nicotinic receptors: implications for drug discovery. Mol Pharmacol 90:288–299
Cottriall CL, Truong HT, McBrien NA (2001) Inhibition of myopia development in chicks using himbacine: a role for M4 receptors. Neuroreport 12:2453–2456
Coulson FR, Fryer AD (2003) Muscarinic acetylcholine receptors and airway diseases. Pharmacol Ther 98(1):59–69
Cragg GM, Newman DJ (2013) Natural products: a continuing source of novel drug leads. Biochim Biophys Acta 1830:3670–3695
Datura (2016) *Datura metel*. plants.ces.ncsu.edu
De Caen AR, Berg MD, Chameides L et al (2015) Part 12: pediatric advanced life support: 2015 American heart association guidelines update for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 132(18):S526–S542
De Maagd GA, Davenport TC (2012) Management of urinary incontinence. P T 37(6):345–361
Douglas CL, Baghdoyan HA, Lydic R (2001) M2 muscarinic autoreceptors modulate acetylcholine release in prefrontal cortex of C57BL/6J mouse. J Pharmacol Exp Ther 299(3):960–966
Duke JA (2002) Handbook of medicinal herbs, 2nd edn. CRC Press, Boca Raton
Eglen RM (2006) Muscarinic receptor subtypes in neuronal and non-neuronal cholinergic function. Auton Autacoid Pharmacol 26(3):219–233
Fink-Jensen A, Fedorova I, Wortwein G, Woldbye DP, Rasmussen T, Thomsen M, Bolwig TG, Knitowski KM, McKinzie DL, Yamada M, Wess J, Basile A (2003) Role for M5 muscarinic acetylcholine receptors in cocaine addiction. J Neurosci Res 74(1):91–96
Foster DJ, Gentry PR, Lizardi-Ortiz JE, Bridges TM, Wood MR, Niswender CM, Sulzer D, Lindsley CW, Xiang Z, Conn PJ (2014) M5 receptor activation produces opposing physiological outcomes in dopamine neurons depending on the receptor's location. J Neurosci 34 (9):3253–3262
Fox C, Richardson K, Maidment ID, Savva GM, Matthews FE, Smithard D, Coulton S, Katona C, Boustani MA, Brayne C (2011) Anticholinergic medication use and cognitive impairment in the older population: the medical research council cognitive function and ageing study. J Am Geriatr Soc 59(8):1477–1483
Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol 63:1256–1272
Garcia-Romeu A, Kersgaard B, Addy PH (2016) Clinical applications of hallucinogens: a review. Exp Clin Psychopharmacol 24(4):229–268
Gautam D, Han SJ, Hamdan FF, Jeon J, Li B, Li JH, Cui Y, Mears D, Lu H, Deng C, Heard T, Wess J (2006) A critical role for (beta cell) M3 muscarinic acetylcholine receptors in regulating insulin release and blood glucose homeostasis in vivo. Cell Metab 3(6):449–461
Gentry PR, Kokubo M, Bridges TM, Cho HP, Smith E, Chase P, Hodder PS, Utley TJ, Rajapakse A, Byers F, Niswender CM, Morrison RD, Daniels JS, Wood MR, Conn PJ, Lindsley CW (2014) Discovery, synthesis and characterization of a highly mAChR selective M5 orthosteric antagonist, VU0488130 (ML381): a novel molecular probe. ChemMedChem 9 (8):1677–1682

Grynkiewicz G, Gadzikowska M (2008) Tropane alkaloids as medicinally useful natural products and their synthetic derivatives as new drugs. Pharmacol Rep 60(4):439–463
Gulenay M, Mathai JK (2020) Depolarizing neuromuscular blocking drugs. Stat Pearls Publishing, Treasure Island
Guo ML, Mao LM, Wang JQ (2010) Modulation of M4 muscarinic acetylcholine receptors by interacting proteins. Neurosci Bull 26(6):469–473
Haga T (2013) Molecular properties of muscarinic acetylcholine receptors. Proc Jpn Acad Ser B Phys Biol Sci 89(6):226–256
Haga K, Kruse AC, Asada H, Yurugi-Kobayashi T, Shiroishi M, Zhang C, Weis WI, Okada T, Kobilka BK, Haga T, Kobayashi T (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature 482(7386):547–551
Hamilton SE, Hardouin SN, Anagnostaras SG, Murphy GG, Richmond KN, Silva AJ, Feigl EO, Nathanson NM (2001) Alteration of cardiovascular and neuronal function in M1 knockout mice. Life Sci 68(22–23):2489–2493
Harvey RD, Belevych AE (2003) Muscarinic regulation of cardiac ion channels. Br J Pharmacol 139:1074–1084
Harvey AR, Baker LB, Reddihough DS, Scheinberg A, Williams K (2018) Trihexyphenidyl for dystonia in cerebral palsy. Cochrane Database Syst Rev 5:CD012430
He Y, Luo J, Kong L (2011) Preparative separation of atropine and scopolamine from Daturae metelis Flos using pH-zone-refining counter-current chromatography with counter-rotation and dual-mode elution procedure. J Sep Sci 34(7):806–811
Ito Y, Kaji M, Sakamoto E, Terauchi Y (2019) The beneficial effects of a muscarinic agonist on pancreatic β-cells. Sci Rep 9(1):16180
IUPAC (2012) Compendium of chemical terminology. Gold book. https://goldbook.iupac.org/. Accessed Mar 2020
Ivanovic SR, Dimitrijevic B, Cupic V, Jezdimirovic M, Borozan S, Savic M, Savic D (2016) Downregulation of nicotinic and muscarinic receptor function in rats after subchronic exposure to diazinon. Toxicol Rep 3:523–530
Jiang S, Li Y, Zhang C, Zhao Y, Bu G, Xu H, Zhang YW (2014) M1 muscarinic acetylcholine receptor in Alzheimer's disease. Neurosci Bull 30(2):295–307
Jo WS, Hossain MA, Park SC (2014) Toxicological profiles of poisonous, edible, and medicinal mushrooms. Mycobiology 42(3):215–220
Katzung B, Trevor A (2014) Basic and clinical pharmacology. McGraw-Hill Education, New York
Kees M, Beckel N, Sharp C (2015) Successful treatment of Solanum dulcamara intoxication in a Labrador retriever puppy. Can Vet J 56(12):1283–1286
Kim YB, Sung TY, Yang HS (2017) Factors that affect the onset of action of non-depolarizing neuromuscular blocking agents. Korean J Anesthesiol 70(5):500–510
Kobayashi F, Yageta Y, Segawa M, Matsuzawa S (2007) Effects of imidafenacin (KRP-197/ONO-8025), a new anti-cholinergic agent, on muscarinic acetylcholine receptors. High affinities for M3 and M1 receptor subtypes and selectivity for urinary bladder over salivary gland. Arzneimittelforschung 57(2):92–100
Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E, Green HF, Liu T, Chae PS, Dror RO, Shaw DE, Weis WI, Wess J, Kobilka BK (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature 482(7386):552–556
Kumar A, Nisha CM, Silakari C, Sharma I, Anusha K, Gupta N, Nair P, Tripathi T, Kumar A (2016) Current and novel therapeutic molecules and targets in Alzheimer's disease. J Formos Med Assoc 115(1):3–10
Lahlou M (2014) The success of natural products in drug discovery. PP 4(3):17–31
Langmead CJ, Watson J, Reavill C (2008) Muscarinic acetylcholine receptors as CNS drug targets. Pharmacol Ther 117:232–243
Lanzafame A, Christopoulos A, Mitchelson F (2001) The allosteric interaction of otenzepad (AF-DX 116) at muscarinic M2 receptors in guinea pig atria. Eur J Pharmacol 416(3):235–244
Le T, Bhushan V (2014) First aid for the USMLE step 1. McGraw-Hill Education, New York

Lebois EP, Thorn C, Edgerton JR, Popiolek M, Xi S (2018) Muscarinic receptor subtype distribution in the central nervous system and relevance to aging and Alzheimer's disease. Neuropharmacology 136:362–373
Lee JH, Francis PT, Ballard CG, Aarsland D, Kalaria RN, Wong PT, Chen CP, Lai MK (2016) Muscarinic M1 receptor coupling to g-protein is intact in Parkinson's disease dementia. J Parkinsons Dis 6(4):733–739
Lieberman JA (2004) Managing anticholinergic side effects. Prim Care Companion J Clin Psychiatry 6(2):20–10
Liu F, Myers AG (2016) Development of a platform for the discovery and practical synthesis of new tetracycline antibiotics. Curr Opin Chem Biol 32:48–57
Macdiarmid SA (2008) Concomitant medications and possible side effects of antimuscarinic agents. Rev Urol 10(2):92–98
Malca Garcia GR, Hennig L, Shelukhina IV, Kudryavtsev DS, Bussmann RW, Tsetlin VI, Giannis A (2015) Curare alkaloids: constituents of a Matis dart poison. J Nat Prod 78:2537–2544
Malik M, Rangel-Barajas C, Sumien N, Su C, Singh M, Chen Z, Huang RQ, Meunier J, Maurice T, Mach RH, Luedtke RR (2015) The effects of sigma (σ1) receptor-selective ligands on muscarinic receptor antagonist-induced cognitive deficits in mice. Br J Pharmacol 172(10):2519–2531
Mathur S, Hoskins C (2017) Drug development: Lessons from nature. Biomed Rep 6(6):612–614
McEvoy GK (2003) Benztropine mesylate: AHFS drug information. American Society of Health-System Pharmacists, Bethesda MD, pp 1185–1186
McEvoy GK (2005) Scopolamine: AHFS drug information. American Society of Health-System Pharmacists, Bethesda MD, pp 1254–1257
McEvoy GK (2018) Atropine: AHFS drug information. American Society of Health-System Pharmacists, Bethesda MD
McFerren SC, Gomelsky A (2015) Treatment of overactive bladder in the elderly female: the case for trospium, oxybutynin, fesoterodine and darifenacin. Drugs Aging 32(10):809–819
Milanlioglu A (2011) Toxic encephalopathy after Atropa belladonna poisoning/ Milanlioglu/Pakistan. Pak J Med Sci 27:26–928
Mintzer J, Burns A (2000) Anticholinergic side-effects of drugs in elderly people. J R Soc Med 93 (9):457–462
Moulton BC, Fryer AD (2011) Muscarinic receptor antagonists, from folklore to pharmacology; finding drugs that actually work in asthma and COPD. Br J Pharmacol 163(1):44–52
Muise ED, Gandotra N, Tackett JJ, Bamdad MC, Cowles RA (2017) Distribution of muscarinic acetylcholine receptor subtypes in the murine small intestine. Life Sci 169:6–10
Nelson ML, Levy SB (2011) The history of the tetracyclines. Ann N Y Acad Sci 1241:17–32
Nesom GL (2012) Notes on the Garrya ovata (Garryaceae) complex. Phyton 97:1–6
Okabe S, Shimosako K, Amagase K (2002) Pharmacological control of gastric acid secretion for the treatment of acid-related peptic disease: past, present, and future. J Physiol Pharmacol 52 (4):639–656
Pergolizzi JV, Philip BK, Leslie JB, Taylor R, Raffa RB (2012) Perspectives on transdermal scopolamine for the treatment of postoperative nausea and vomiting. J Clin Anesth 24:334–345
Prado VF, Janickova H, Al-Onaizi MA, Prado MAM (2017) Cholinergic circuits in cognitive flexibility. Neuroscience 345:130–141
Prommer E (2013) Anticholinergics in palliative medicine: an update. Am J Hosp Palliat Care 30 (5):490–498
Qin K, Dong C, Wu G, Lambert NA (2011) Inactive-state preassembly of Gq- coupled receptors and Gq heterotrimers. Nat Chem Biol 7(10):740–747
Rajput H (2013) Effects of atropa belladonna as an anti-cholinergic. Nat Prod Chem Res 1:104. https://doi.org/10.4172/2329-6836.1000104
Ramnarine M (2020) Anticholinergic toxicity. https://emedicine.medscape.com/article/812644-overview. Accessed Mar 2020
Rang HP (2003) Pharmacology, 5th edn. Churchill Livingstone, Edinburgh

Roberts MF, Wink M (1999) Alkaloids: biochemistry, ecology and medicinal applications. J Nat Prod 62(4):662–664

Robinson L, Platt B, Riedel G (2011) Involvement of the cholinergic system in conditioning and perceptual memory. Behav Brain Res 221:443–465

Sanchez G, Colettis N, Vazquez P, Cervenansky C, Aguirre A, Quillfeldt JA, Jerusalinsky D, Kornisiuk E (2009) Muscarinic inhibition of hippocampal and striatal adenylyl cyclase is mainly due to the M(4) receptor. Neurochem Res 34(8):1363–1371

Saternos HC, Almarghalani DA, Gibson HM, Meqdad MA, Antypas RB, Lingireddy A, AbouAlaiwi WA (2018) Distribution and function of the muscarinic receptor subtypes in the cardiovascular system. Physiol Genomics 50(1):1–9

Schiechl G, Himmelsbach M, Buchberger W, Kerschbaum HH, Lutz-Meindl U (2008) Identification of acetylcholine and impact of cholinomimetic drugs on cell differentiation and growth in the unicellular green alga Micrasterias denticulate. Plant Sci 175(3):262–266

Sheela JRH (2013) Qualitative analysis of secondary metabolites of the plant clematis gouriana. Int J Innov Res Sci Eng Technol 2(6):2356–2358

Snyder PJ, Bednar MM, Cromer JR, Maruff P (2005) Reversal of scopolamine-induced deficits with a single dose of donepezil, an acetylcholinesterase inhibitor. Alzheimers Dement 1 (2):126–135

Stepnicki P, Kondej M, Kaczor AA (2018) Current concepts and treatments of schizophrenia. Molecules 23(8):2087

Svoboda J, Popelikova A, Stuchlik A (2017) Drugs interfering with muscarinic acetylcholine receptors and their effects on place navigation. Front Psych 8:215

Thal DM, Sun B, Feng D, Nawaratne V, Leach K, Felder CC, Bures MG, Evans DA, Weis WI, Bachhawat P, Kobilka TS, Sexton PM, Kobilka BK, Christopoulos A (2016) Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. Nature 531(7594):335–340

Tolaymat M, Larabee SM, Hu S, Xie G, Raufman JP (2019) The role of M3 muscarinic receptor ligand-induced kinase signaling in colon cancer progression. Cancers 11:308

Toxic plants (2020) Safe and poisonous garden plants. University of California. https://ucanr.edu/sites/poisonous_safe_plants/Toxic_Plants_by_common_Name_659/. Accessed Mar 2020

Tytgat GN (2007) Hyoscine butylbromide: a review of its use in the treatment of abdominal cramping and pain. Drugs 67(9):1343–1357

Tzavara ET, Bymaster FP, Davis RJ, Wade MR, Perry KW, Wess J, McKinzie DL, Felder C, Nomikos GG (2004) M4 muscarinic receptors regulate the dynamics of cholinergic and dopaminergic neurotransmission: relevance to the pathophysiology and treatment of related CNS pathologies. FASEB J 18(12):1410–1412

Ulbricht C, Basch E, Hammerness P, Vora M, Wylie J Jr, Woods J (2004) An evidence-based systematic review of belladonna by the natural standard research collaboration. J Herb Pharmacother 4(4):61–90

Valecha N, Looareesuwan S, Martensson A, Abdulla SM, Krudsood S, Tangpukdee N, Mohanty S, Mishra SK, Tyagi PK, Sharma SK, Moehrle J, Gautam A, Roy A, Paliwal JK, Kothari M, Saha N, Dash AP, Bjorkman A (2010) Arterolane, a new synthetic trioxolane for treatment of uncomplicated Plasmodium falciparum malaria: a phase II, multicentre, randomized, dose-finding clinical trial. Clin Infect Dis 51(6):684–691

Van Rhee AM, Jacobson KA (1996) Molecular architecture of G protein-coupled receptors. Drug Dev Res 37:1–38

Verma S, Kumar A, Tripathi T, Kumar A (2018) Muscarinic and nicotinic acetylcholine receptor agonists: current scenario in Alzheimer's disease therapy. J Pharm Pharmacol 70(8):985–993

Vuckovic Z, Gentry PR, Berizzi AE, Hirata K, Varghese S, Thompson G, van der Westhuizen ET, Burger WAC, Rahmani R, Valant C, Langmead CJ, Lindsley CW, Baell JB, Tobin AB, Sexton PM, Christopoulos A, Thal DM (2019) Crystal structure of the M5 muscarinic acetylcholine receptor. Proc Natl Acad Sci U S A 116(51):26001–26007

Wang H, Peng G, Wang B, Yin H, Fang X, He F, Zhao D, Liu Q, Shi L (2020) IL-1R−/− alleviates cognitive deficits through microglial M2 polarization in AD mice. Brain Res Bull 157:10–17

Wawruch M, Macugova A, Kostkova L, Luha J, Dukat A, Murin J, Drobna V, Wilton L, Kuzelova M (2012) The use of medications with anticholinergic properties and risk factors for their use in hospitalised elderly patients. Pharmacoepidemiol Drug Saf 21(2):170–176

Wess J, Maggio R, Palmer JR, Vogel Z (1992) Role of conserved threonine and tyrosine residues in acetylcholine binding and muscarinic receptor activation. A study with m3 muscarinic receptor point mutants. J Biol Chem 267(27):19313–19319

Wess J, Eglen RM, Gautam D (2007) Muscarinic acetylcholine receptors: mutant mice provide new insights for drug development. Nat Rev Drug Discov 6:721–733

Weston GK, Huang XF, Lian J, Deng C (2012) Effects of olanzapine on muscarinic M3 receptor binding density in the brain relates to weight gain, plasma insulin and metabolic hormone levels. Eur Neuropsychopharmacol 22(5):364–373

Witkin JM, Overshiner C, Li X, Catlow JT, Wishart GN, Schober DA et al (2014) M1 and M2 muscarinic receptor subtypes regulate antidepressant-like effects of the rapidly acting antidepressant scopolamine. J Pharmacol Exp Ther 351:448–456

Xu D, Anderson HD, Tao A, Hannah KL, Linnebur SA, Valuck RJ, Culbertson VL (2017) Assessing and predicting drug-induced anticholinergic risks: an integrated computational approach. Ther Adv Drug Saf 8(11):361–370

Yadav KN, Kadam PV, Patel JA, Patil MJ (2014) Strychnos Potatorum: phytochemical and pharmacological review. Pharmacogn Rev 8(15):61–66

Yazdani N, Sadeghi R, Momeni-Moghaddam H, Zarifmahmoudi L, Ehsaei A (2018) Comparison of cyclopentolate versus tropicamidecycloplegia: a systematic review and meta-analysis. J Optom 11(3):135–143

Dopamine Beta Hydroxylase: An Enzyme with Therapeutic Potential to Combat Neural and Cardiovascular Diseases

14

Swati Kundu, Manisha Saini, Sanjay Kumar Dey, and Suman Kundu

Abstract

The brain and the heart are arguably the two most important organs of the human body. It is thus no surprise that diseases of the brain and heart are of the highest concern and are the major causes of mortality and morbidity worldwide. A physiological process that is common to both of these major organs is the catecholamine biosynthetic pathway, where the products of the pathway regulate several major events in the human body. The changes in the levels of catecholamines are originators of several neural and cardiovascular diseases. Dopamine beta hydroxylase (DBH), an enzyme that plays a central and critical role in the catecholamine biosynthetic pathway, regulates the concentrations of dopamine and norepinephrine, whose deficiency or overproduction causes several diseases related to the brain and the heart. This enzyme is thus of great therapeutic significance. Insight into the genetics, structure, function, and dynamics of the protein will provide scope for discovery and design of potential small molecule drugs to treat neurological or cardiovascular disorders utilizing structure-based, rational drug discovery approaches.

Keywords

Dopamine beta hydroxylase · Catecholamine · Dopamine · Norepinephrine · Hypertension · Depression · Cocaine addiction

Swati Kundu and Manisha Saini contributed equally with all other contributors.

S. Kundu · M. Saini · S. K. Dey · S. Kundu (✉)
Department of Biochemistry, University of Delhi, New Delhi, India
e-mail: suman.kundu@south.du.ac.in

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_14

14.1 Introduction to DBH

DBH (EC 1.14.17.1) is a copper-dependent monooxygenase, which has broad substrate specificity, and oxidizes any phenylethylamine to its corresponding phenylethanolamine (Levin 1961; Tishchenko et al. 2016). One of the specific examples is catalysis of dopamine into norepinephrine (and thus epinephrine) in mammalian tissues and serum (Kruse et al. 1987; Tishchenko et al. 2016). As shown in Fig. 14.1, the reaction is catalyzed in the presence of other cofactors such as molecular oxygen and ascorbic acid where DBH acts as an oxidoreductase enzyme that inserts one oxygen atom from molecular oxygen into the substrate, whereas ascorbic acid serves as the reductant for the reaction (Rush and Geffen 1980; Ross and Rosenzweig 2017). The reaction starts with the hydroxylation of the substrate (dopamine) whereby molecular oxygen donates single oxygen atom and ascorbic acid acts as an electron donor (Ljones and Flatmark 1974; Beliaev et al. 2009). A ping pong model is suggested for substrate binding in which the enzymic copper in cupric state gets reduced to a cuprous state, which binds to oxygen and thereby hydroxylates the substrate (Beliaev et al. 2009). During this reaction, the formation of dehydroascorbate occurs from ascorbate. This reaction is a key step in the catecholamine biosynthetic pathway that regulates several physiological processes which are controlled by the brain like movement, pleasure, attention, mood, and motivation. The reaction also regulates processes under the control of peripheral nervous system like maintenance of blood pressure (BP), dilation of pupils and bronchioles, increasing renin secretion from the kidneys and inhibiting peristalsis (William Tank and Lee Wong 2011).

14.1.1 Cellular Distribution of DBH

DBH is incorporated in the neurons where it is stored in the storage vesicles of the central nervous system, the sympathetic ganglia, while in the periphery it is present in chromaffin granules of the adrenal medulla (Teitelman et al. 1979; Kemper et al. 1987; Catelas et al. 2020). As the concentration of noradrenergic cells is higher in the

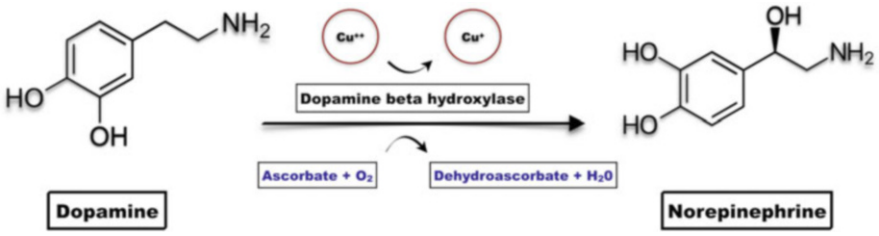


Fig. 14.1 Reaction catalyzed by DBH, an oxidoreductase belonging to copper type II, ascorbate-dependent monooxygenase

brain, DBH becomes a prime marker for these cells (Sokoloff et al. 1985; Lewis and Asnani 1992; Punchaichira et al. 2018). In the brain, noradrenergic neurons in the nerve terminals contain catecholamine vesicles inside which DBH is synthesized (Biaggioni et al. 1990). In the adrenal medulla, DBH is present in the membrane-bound and soluble forms in which the latter is derived from the membrane-bound isoforms (Laduron 1975; Sokoloff et al. 1985). Both forms have similar tetrameric structure, immunoreactivity, carbohydrate contents, and kinetic parameters (Wallace et al. 1973; Slater et al. 1981; Saxena and Fleming 1983; Kapoor et al. 2011). It was demonstrated that non-covalently bound phosphatidyl serine moieties are involved in the anchoring of membrane DBH to chromaffin granules (Taylor and Fleming 1989).

In the central nervous system, DBH is associated with neurocognitive function such as working memory, behavioral traits including impairments in active-avoidance learning, memory retrieval, and maternal and social behavior (Marino et al. 2005; Parasuraman et al. 2005; Robbins and Arnsten 2009). In the peripheral system (DBH in adrenal medulla), the enzyme is involved in BP regulation and altered cardiac function (Rapacciuolo et al. 2001; Whaley-Connell et al. 2006).

14.1.2 Storage and Release of DBH

DBH is associated with highly concentrated monoamine neurotransmitters in the storage vesicles as shown in Figs. 14.2 and 14.3. Upon neuronal activation, calcium-dependent exocytosis of DBH along with the neurotransmitters occurs as vesicle contents are released from the nerve terminals into the extracellular space (Fig. 14.2) (Del-Bel and De-Miguel 2018). Since there is no classic postsynaptic specialization associated with the majority of nerve endings, released contents diffuse to postsynaptic cells in the vicinity where they stimulate their respective receptors (Lodish et al. 2000).

Exocytosis of catecholamines from chromaffin cells and sympathetic axons also occurs (Fig. 14.3). At both the locations, degranulation occurs, and components enter into the circulation (Del-Bel and De-Miguel 2018). As a result of exocytosis, DBH is also released from these vesicles into extracellular space and thus can be found in plasma and cerebrospinal fluid (De Potter et al. 1970; Johnson et al. 1971; Weinshilboum et al. 1971; Catelas et al. 2020). The concentrations of DBH in the blood vary widely in the general population (1–100 U/L); however, the interindividual variation in DBH activity is low (0.8–21.3 U/L) (Deinum et al. 2004; Nagatsu 2009). For the catalysis of reaction, the substrate of DBH (dopamine) is taken up from the cytosol to vesicles where it is converted to norepinephrine.

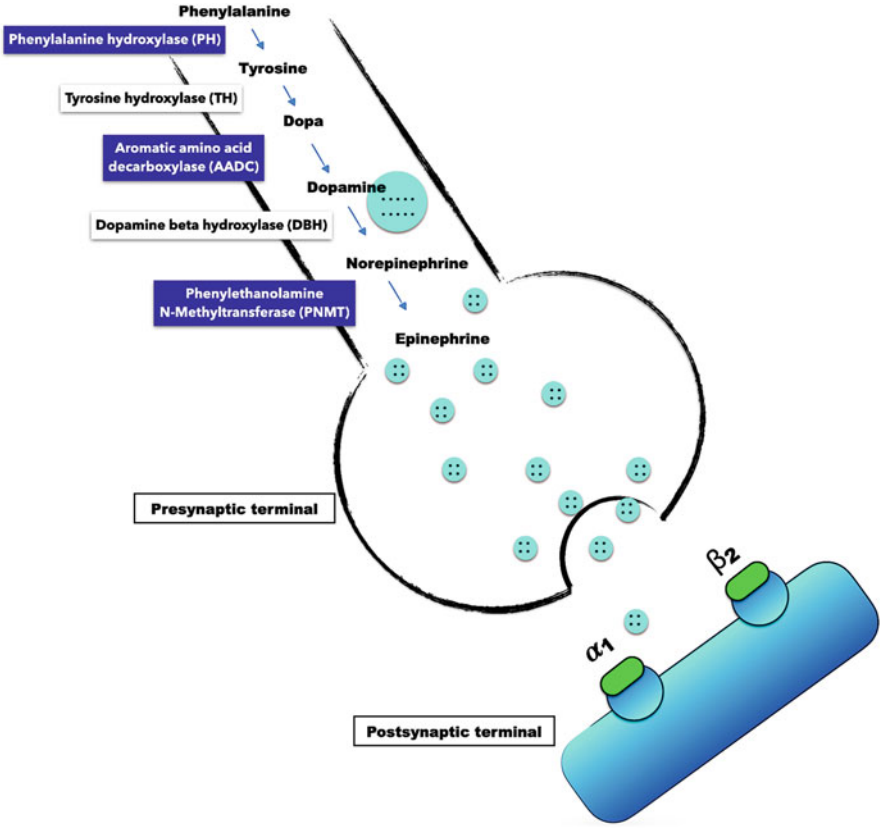


Fig. 14.2 Storage, synthesis, and release of neurotransmitters (catecholamines) and DBH in sympathetic nerve ganglion

14.2 Structure of DBH and Its Therapeutic Implications

14.2.1 Genetic Organization

DBH gene is 23 kb long with 12 exons situated on chromosome 9 (Fig. 14.4). Linkage analysis by Goldin et al. (1982) inferred that DBH activity is mediated by a single quantitative trait locus in a region on chromosome 9 (9q34) (Goldin et al. 1982). A complete gene map of *DBH* is available (Cubells et al. 2000). Studies based on sequencing analyses identified the molecular structure of *DBH* and an SNP in the promoter region-rs1611115/C-970 T/formerly C-1021 T (Zabetian et al. 2001) was reasoned to be the source for the wide interindividual variation in DBH activity in plasma.

A few mutations in *DBH* which result in the norepinephrine deficiency has also been documented (Kim et al. 2011; Garland 2012). Change in alleles, T (thymine) of

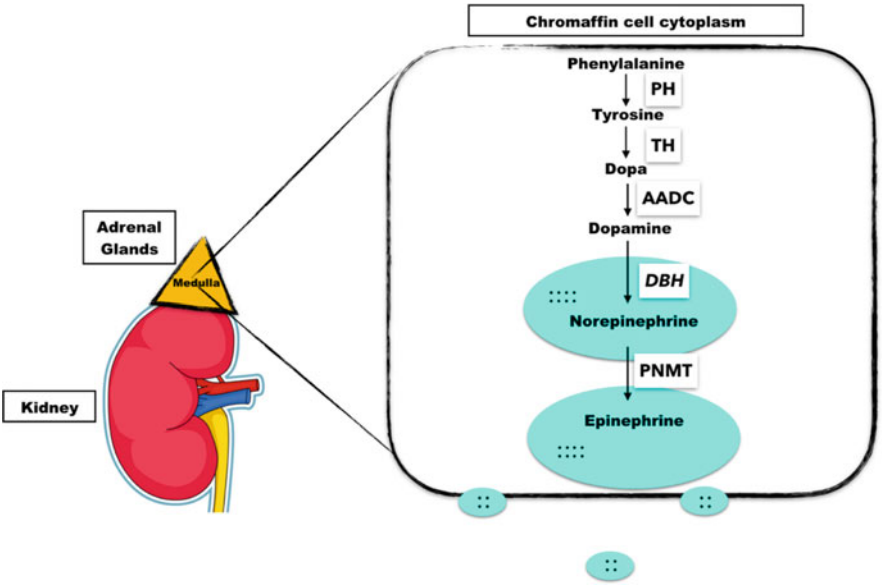


Fig. 14.3 Storage, synthesis, and release of neurotransmitters (catecholamines) and DBH in chromaffin cells of the adrenal medulla

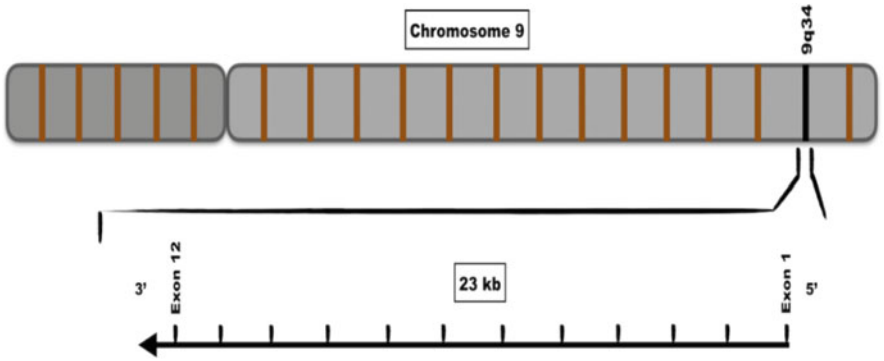


Fig. 14.4 Schematic of DBH gene located on the q-arm of chromosome 9

the −1021 C/T SNP and A (adenine) allele of the 444 G/A SNP is associated with lower DBH enzyme activity (Parasuraman et al. 2012). Norepinephrine deficiency due to lower DBH activity is directly linked to diseases such as attention-deficit hyperactivity disorder (ADHD), depression, and hypotension (low BP) (Garland 2012), thereby necessitating the discovery of activators of DBH enzyme. Similarly, some mutations in DBH were shown to result in their altered activity, which has been reported in neurological disorders and hypertension (Matuzas et al. 1982; Abe et al. 2005). It is proposed that such neurological disorders as well as hypertension

Table 14.1 DBH inhibitors and their shortcomings

S. no.	DBH inhibitors	Drawbacks	References
1.	4-Aminooxymethylphenol and 3-(*N*-methylhydrazinomethyl) phenol (NSD 1034), disulfiram, and FLA 63	Chelates copper from any copper containing metalloenzyme, nonspecific.	Lippmann and Lloyd (1969); Corrodi et al. (1970)
2.	1-Phenyl-3-thiazol-2-ylthiourea (U-14,624), 2-[2-benzimidazolyl]-amino-2-imidazoline dihydrochloride (BRL 8242) and fusaric acid	Cause orthostatic hypotension; increases the heart rate, nausea, and palpitations	Johnson et al. (1970); Claxton et al. (1976)
3.	Imidazole-2-thione-based inhibitors: (nepicastat, etamicastat, and zamicastat)	Nepicastat crosses the blood–brain barrier	Kruse et al. (1986); Rocha et al. (2012)

may be treated by using DBH inhibitors such as nepicastat, etamicastat, and zamicastat (Table 14.1). This aspect has been detailed in a recent review article (Dey et al. 2020).

Additive effects of SNPs in the *DBH* gene were also reported. Thus, individuals carrying two copies of the T allele of the −1021 C/T SNP (TT) and two copies of the A allele of the 444 G/A SNP (AA) demonstrated decreased activity of DBH (Parasuraman et al. 2012). Five putatively functional non-synonymous variants, namely L317P, A318S, D460N, W544S, and R549C were characterized by our group recently which provided a new insight into the structure–function relationship of DBH (Punchaichira et al. 2017). It is thus evident that genetic defects in DBH are debilitating and provide scope for therapeutic interventions.

14.2.2 Molecular Properties of DBH Enzyme

DBH protein is glycosylated with a tetrameric structure of 290 kDa (Frigon and Stone 1978). Each monomer of 68 kDa consists of 1 sialic acid, 1 glucose, 2 galactose, 5 *N*-acetyl glucosamine, and 7 mannose molecules covalently attached (Vendelboe et al. 2016). The human DBH gene encodes a 617-amino-acid protein with two copper binding domains at the active site—CuM and CuH. The C-terminus, known as the dimerization domain, consists of 100 amino acids. Of the 15 cysteine residues present in DBH, 14 cysteines are involved in disulfide bridge formation, of which six are intramolecular bonds and two are intermolecular bonds (Vendelboe et al. 2016).

The crystal structure of DBH is in dimeric form (Vendelboe et al. 2016), which is an inactive form (Fig. 14.5). An in silico model proposed prior to the crystal structure is of tetrameric form that shares significant structural homology with the crystal structure in the active site domain (Kapoor et al. 2011). The tetrameric enzyme was shown to be the active form of the enzyme in the past (Houhou et al. 1995). The in silico model predicted that each dimer forms intramolecular disulfide linkage

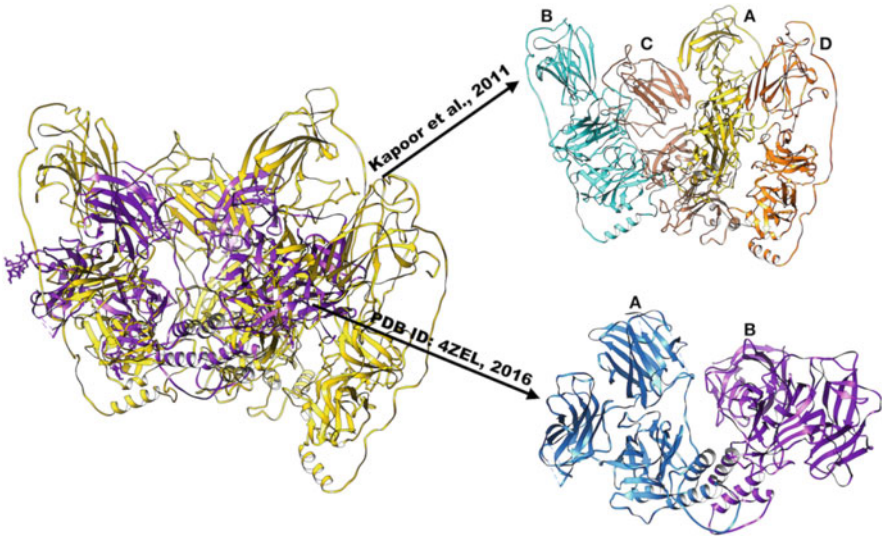


Fig. 14.5 Dimeric crystal structure of DBH (PDB ID: 4ZEL) aligned with tetrameric in silico structure of DBH. The active sites in both the models are similar. The alphabets (A, B, C, and D) indicate the various subunits

between cysteines at positions 154–596, 232–283, 269–295, 390–503, 394–565, and 466–488, whereas two cysteines at positions 528 and 530 form intermolecular disulfide bonds, thereby resulting in tetrameric structure (Kapoor et al. 2011). CuH is coordinated to three histidines (His262, His263, and His333), whereas CuM to two histidines (His412 and His414) and Met487 (Vendelboe et al. 2016) (Fig. 14.6). The three histidines interact with copper through their delta nitrogen atom at the CuM binding site. Similarly, histidines interact with copper through their epsilon nitrogen atom at the CuH binding site. The distance of CuM is 1.96 Å and 1.97 Å, and of CuM is 2.08 Å and 2.02 Å from the copper center, respectively. DBH has three domains, namely DOMON domain, Cu type II N-terminal domain, and C-terminal domain. All these features have been outlined in detail in the in silico (except Met487) and in the crystal structure (Kapoor et al. 2011; Vendelboe et al. 2016). There are two conformations of the enzyme—closed and open. In the closed conformation, the copper site is catalytically active, and in the open conformation, loading of substrate and release of product is described. The three-dimensional structure of the DBH protein will allow mapping of the SNPs described above and design of activators or inhibitors of DBH for therapeutic intervention.

14.2.3 Assay Methods

The enzyme activity of DBH may be measured either by monitoring the depletion of the substrate dopamine or by monitoring the increase in concentration of

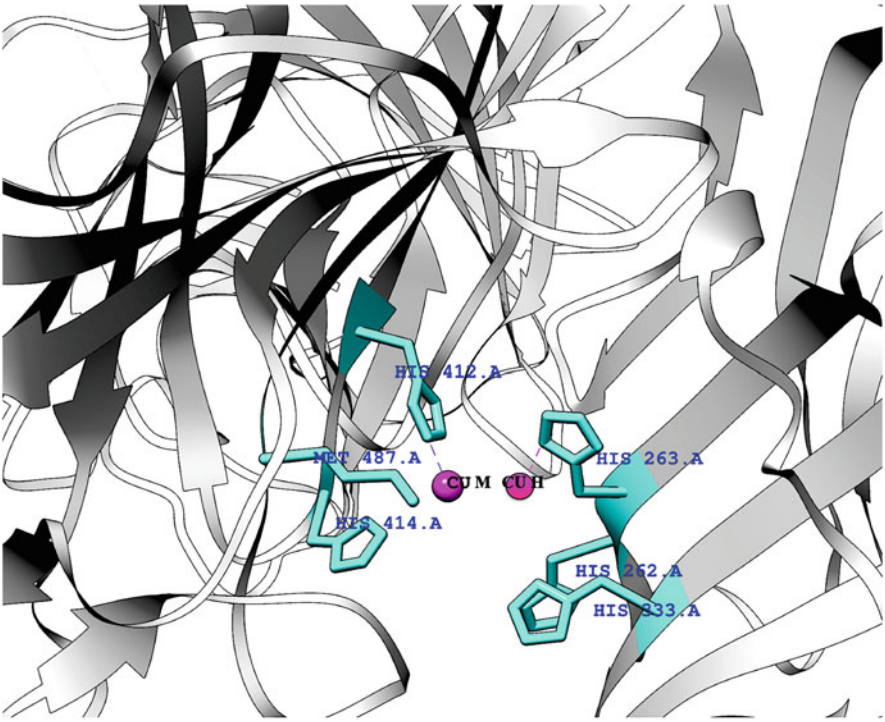


Fig. 14.6 Binding pocket of DBH comprising of His262, His263, His333, His412, His414, and Met487 in interaction with the copper atoms CuH and CuM, respectively

norepinephrine over the assay period. Due to higher stability of tyramine, however, the assay frequently follows the conversion of tyramine to octopamine, instead of dopamine to norepinephrine (Molinoff et al. 1969). One of the earliest developments was a radioactivity-based assay using phenylethanolamine-*N*-methyltransferase that got success in measuring DBH activity for the first time in 1971 (Johnson et al. 1971; Weinshilboum et al. 1971). This assay depends upon the β-hydroxylation of tyramine to form octopamine, which is enzymatically N-methylated by phenylethanolamine-*N*-methyltransferase, with a C-labeled methyl group donated by *S*-adenosyl-methionine-methyl-C. Another assay utilized the same compound (tyramine as a substrate), but the activity was measured by photometry (Nagatsu 2009). Recently, a chromatographic method for assaying DBH activity was reported (Punchaichira et al. 2018), where the reaction mixture consisting of the enzyme and the substrate tyramine was subjected to ultra-performance liquid chromatography (UPLC), thus separating the product formed, octopamine, from any unutilized substrate. The elution time for substrate and product was measured and their relative absorbance was used to quantitate the conversion. The UPLC method is a robust assay but utilized expensive reagents. Cheaper and high-throughput assays need to be developed for the purposes of drug discovery with DBH as a target.

Moreover, activity assays for DBH are quite difficult to perform in untreated samples including blood as it has natural quenchers such as sulfhydryl compounds which bind to copper and thereby inhibit the enzyme (Nagatsu 2009). However, this problem can be solved by treating these inhibitors with chemicals such as an SH-blocking agent like *N*-ethylmaleimide (Nagatsu et al. 1967). The photometric assay described above also used *N*-ethylmaleimide that act against these inhibitors present in human serum.

14.3 Factors That Affect DBH Gene and Its Activity

DBH, the locus encoding DBH protein, is the major quantitative trait locus controlling circulating DBH levels (Cubells et al. 2000). Genetic factors are known to cause individual variation in the levels of DBH protein and its activity; however, the stability of the enzyme over a long time in normal humans makes it a clinically significant enzyme (Weinshilboum 1989; Punchaichira et al. 2017). DBH gene and its two forms of mRNAs, namely type A and type B, are regulated by transcriptional regulatory elements including glucocorticoid and cAMP-responsive elements which are placed near the transcription start site of DBH gene (Kobayashi et al. 1989). The expression of DBH in various cells is known to be mediated by multiple nuclear factors in the upstream promoter that surrounds the cAMP response element (Seo et al. 1996; Stubbusch et al. 2011). Several studies showed that the major determinant in DBH release is neuronal stimulation through the alpha pre-synaptic receptors (Johnson et al. 1971; Weinshilboum et al. 1971; Catelas et al. 2020). Some inhibitors used against brain DBH were found not to alter the enzyme activity but the rate of synthesis of DBH was changed (Rosenberg and Lovenberg 1977). However, disulfiram and nepicastat are known to reduce DBH activity which leads to decreased norepinephrine levels and increased tissue dopamine levels (Schroeder et al. 2010; De La Garza et al. 2015) with significant therapeutic implications like lowering of BP or treatment of Parkinson's disease. The other factor is DBH itself since a significant amount of the enzyme is leaked out of neurons which may go into CSF from blood or brain cells (Hastings et al. 2004). Such leakage is the reason behind a higher concentration of DBH in CSF than in plasma in humans (Lerner et al. 1978). However, the mechanism of the release and reduced DBH activity are still not well understood. Mechanistic understanding of DBH production, release, and decreased enzymatic activity are the need of the hour for a successful therapeutic application of the protein.

14.4 Etiology

A variety of disorders due to DBH-related deficiency came into light in 1986 when congenital syndrome was characterized in Netherland (Veld et al. 1987). The findings suggested that low norepinephrine and epinephrine levels and their metabolites in the presence of elevated dopamine concentrations caused orthostatic

hypotension, noradrenergic failure, impairment in cardiovascular autonomic regulation, and insulin resistance (Robertson et al. 1986; Jepma et al. 2011; Arnold et al. 2017). It was later clinically demonstrated that it was due to congenital deficiency of DBH. Presyncopal symptoms in these patients have included dizziness, blurred vision, dyspnea, nuchal discomfort, and occasional chest pain. The effects of this deficiency were also observed in the next generation as affected mothers gave stillbirth or sick births or other clinical abnormalities (Biaggioni et al. 1990). These children were generally hypotensive and hypoglycemic and therefore had a markedly low physical stamina (Mathias et al. 1990; Robertson et al. 1986). These symptoms worsened in adolescence and adulthood (Biaggioni et al. 1990).

The concentrations of endogenous catecholamines like norepinephrine, epinephrine, and dopamine are the most essential determinants in the maintenance of vascular tone and regulation of arterial pressure (Byrne et al. 2018). Their level affects the sympathetic and parasympathetic activities in humans which regulate autonomic functions such as cardiac and vascular tones (McCorry 2007). It is therefore clear that the rise in norepinephrine levels causes an increase in BP in the periphery; however, it causes a reduction in BP within the central nervous system (William Tank and Lee Wong 2011). The other catecholamine/hormone, dopamine, acts as a precursor to both epinephrine and norepinephrine and thereby regulates BP through Class C L-type calcium channels (Shen et al. 2008). This points toward the importance of the regulation of synthesis of norepinephrine/dopamine in controlling BP. The enzyme DBH, which converts dopamine into norepinephrine, is thus of great biochemical and pharmacological importance.

The level of DBH in blood, lymph, and CSF indicates the controlled and regulated activity of noradrenergic nerves (Hammerschmidt et al. 2013). However, the difficulty in measuring DBH activity in CSF, wide variation in resting blood DBH level in humans, and the influence of genetic/environmental factors on its activity limit the possibility of using DBH as a reliable parameter for diagnosis. Therapeutic potential of DBH though, as a target to combat various diseases, seems to be a distinct possibility.

14.5 Therapeutic Importance of DBH

DBH, as evident above, through its influence on the levels of catecholamines, could adversely affect several physiological processes and thus associated with a multitude of diseases. These diseases could either be related to the CNS or the peripheral system. Thus, DBH is associated with diseases like depression, ADHD, migraine, Parkinson's, Alzheimer's disease, schizophrenia, congenital syndromes as well as hypo- or hypertension, hypoglycemia, and ptosis (Tang et al. 2018). An approach to alter sympathetic function is by reducing norepinephrine synthesis. Similarly, peripheral DBH can be inhibited in the adrenal medulla to regulate its effects on visceral organs including cardiovascular system.

Norepinephrine is an important determinant in the neural regulation of BP, while dopamine plays an important depressor role via renal and plasma volume mechanisms. One special significance of DBH is its connection to cardiovascular diseases. Alterations in the DBH activity and/or genetic variants in the DBH gene lead to changes in the sympathoadrenal activity, which is implicated in the pathogenesis of hypertension and cardiovascular disease (Chen et al. 2010; Punchaichira et al. 2017). This makes DBH of high interest to the cardiovascular field. A number of inhibitors of DBH have been designed to combat cardiovascular diseases. Any of these inhibitors that cross the blood–brain barrier could however play a significant role in treating brain-related disorders like cocaine addiction and Parkinson's diseases.

14.5.1 Inhibitors of DBH: Hypertension and Cocaine Addiction

The diseases/disorders of the heart and brain that have been the prime targets for therapeutic intervention via DBH are hypertension and cocaine addiction. In humans, the pathogenesis of hypertension, both primary (genetic, essential) and secondary (acquired), is attributed to an exaggerated activity of the sympathoadrenal machinery (De et al. 1993). Inhibition of DBH enzyme has been found to be of great utility in treating hypertension in spontaneously hypertensive rats (Beliaev et al. 2009; Igreja et al. 2016; Ohlstein et al. 1987; Kruse et al. 1987; Pires et al. 2015). Plasma DBH activity of converting dopamine into norepinephrine ranged from 2 to 100 U/L of plasma in humans (Beliaev et al. 2009). Increased activity of DBH may result in the accumulation of norepinephrine in the chromaffin cells of the adrenal medulla, and therefore to lower dopamine in contrast to norepinephrine levels; conversely, decreased DBH activity is linked to high dopamine levels compared to norepinephrine (Cubells et al. 2000). Given a link between these catecholamines and BP, DBH is expected to be associated with the control of BP. There is considerable activity globally to design inhibitors of DBH to combat hypertension.

Similarly, inhibition of DBH in the central nervous system is being actively considered to combat brain-related disorders like cocaine addiction (Bicker et al. 2016; Loureiro and Soares-da-Silva 2015). Disulfiram (https://clinicaltrials.gov/ct2/show/NCT00395850) and nepicastat (https://clinicaltrials.gov/ct2/show/NCT01704196) have been averred to be beneficial in treating cocaine addiction (Schroeder et al. 2010; De La Garza et al. 2015). The effectiveness of these molecules to cure cocaine enslavement can be achieved by a high amount of brain dopamine caused by the inhibition of DBH. This effect is therapeutic for cocaine dependence since an increase in brain dopamine helps in diminishing withdrawal syndromes and craving, while at the same time a decrease in norepinephrine aids in attenuating relapse to drug use. On the other hand, it has been argued that increase in dopamine release in case of cocaine addiction results in anxiety and dysphoria, not in euphoric response, and thereby causing reduced cocaine use (Kosten et al. 2013; De La Garza et al. 2015).

14.5.2 Current Studies on DBH Inhibitors

In plasma, the activity of DBH was investigated, and multiple studies are documented on the same. Many reports suggest that DBH inhibition may allow the regulation of BP and hence may help to combat hypertension (Barrie et al. 2014; Nunes et al. 2010). A large number of DBH inhibitors are known which reduces norepinephrine both peripherally and centrally; however, not all of them limit BP. In 1976, fusaric acid was suggested as a potential DBH inhibitor in hypertensive and normotensive patients (Furuta and Washizaki 1976). Among the existing DBH inhibitors, disulfiram (a copper chelator) is a nonspecific inhibitor of DBH and is nonresponsive in specific populations (Beliaev et al. 2009). Nepicastat is a potent and specific inhibitor of DBH, but it crosses blood–brain barrier with limited scope against hypertension (Rocha et al. 2012). Nepicastat and disulfiram, however, show potential for brain-related disorders including cocaine addiction (Bicker et al. 2016; Loureiro and Soares-da-Silva 2015; Loureiro et al. 2015). The other inhibitors, etamicastat and zamicastat, which have undergone clinical trials in hypertensive patients, do not cross the blood–brain barrier and are likely to be developed as antihypertensives (Almeida et al. 2013; Igreja et al. 2016). Due to the limited number of useful DBH inhibitors (Table 14.1), there is a need to focus on the identification of new, potent, and specific DBH inhibitors. Table 14.2 summarizes the key inhibitors of DBH that are under clinical trials or otherwise.

An important aspect that needs to be considered for the discovery of a potential DBH inhibitor is whether it is peripherally active (for treating hypertension and other cardiovascular diseases) or it crosses the blood–brain barrier (for treating brain-related disorders). A new class of inhibitors have been identified and validated by our group using structure-based drug discovery methods to combat hypertension and other cardiovascular diseases (Indian patents 201711036983A and 201811005899A) (Kundu et al. 2017, 2018).

14.6 Future Perspectives

The investigation of this classical target, DBH, is of great potential since it provides us with a model that may help us to determine its role in controlling cocaine addiction, ADHD and other neural diseases, systemic hypertension, pulmonary hypertension, and cardiac hypertrophy in humans. This indicates pharmacological and market values of DBH, which needs to be investigated on priority. Studies on DBH as a pharmacological target must be accelerated by focusing on the understanding of the regulation of its release, reaction, and active states present in humans. Emphasis must be on identifying novel inhibitors of DBH. Among the existing inhibitors of DBH, nepicastat appears to be the most promising drug for cocaine dependence, while etamicastat and zamicastat for hypertension (Almeida et al. 2013; Igreja et al. 2019a, b; Schroeder et al. 2013). An advantage of etamicastat over the other inhibitors in treating hypertension is that it does not cross blood–brain barrier.

Table 14.2 Key DBH inhibitors and their current status

S. no.	Inhibitors	Screened in species	Current status	Target disease	References
1.	Disulfiram	Rats and humans	FDA-approved drug in market	Alcohol dependence and Antabuse, cocaine addiction	Hamblin et al. (2019); Kiriakov et al. (1973)
2.	Fusaric acid	Mice, rats, and humans	Not in commercial use	Hypertension	Hachisu et al. (1983); Matsuzaki et al. (1976)
3.	5-(4′-Chlorobutyl) picolinic acid	SHR and dogs	Not in commercial use	Hypertension	Wang and Ng (1999); Ishii et al. (1975)
4.	Nepicastat	SHR, dogs and humans	In clinical trials for cocaine dependence	Cocaine dependence, hypertension	Bonifácio et al. (2015); Sabbah et al. (2000)
5.	Etamicastat	SHR, dogs, monkeys, humans	Phase I (completed) Phase II (about to start)	Hypertension	Almeida et al. (2013); Loureiro and Soares-da-Silva (2015); Stewart et al. (2018)
6.	Zamicastat	SHR, SS rat, humans	Phase I (ongoing)	Cardiac hypertrophy, pulmonary arterial hypertension, and congestive heart failure	Igreja et al. (2016)

Zamicastat is being investigated in details as an alternative (under clinical trials) (Bonifácio et al. 2015; Rocha et al. 2012). Nepicastat, targeting DBH activities in the CNS to treat cocaine addictions and post-traumatic stress disorders, is also in the active clinical trials (https://clinicaltrials.gov/ct2/show/NCT00659230 and https://clinicaltrials.gov/ct2/show/NCT01704196) (De La Garza et al. 2015). It is also important to work out cheaper routes of chemical synthesis for the inhibitors of DBH to finally commercialize the compounds at an affordable cost. A major lacuna is lack of efforts in identifying activators of DBH since it is evident that several diseases may be treated by augmenting the activity of deficient DBH. Computer-aided, structure-based, rational drug discovery approaches are yet to be employed in the discovery of activators or inhibitors of DBH.

14.7 Conclusion

This chapter summarizes the present state of knowledge concerning the molecular properties, cellular localization, genetic and physiological regulation of DBH. It further focuses on the importance of analyzing the structural and functional dynamics of DBH to treat neural or peripheral disorders exploiting its neural or peripheral counterparts, respectively. Recent developments in the later directions have also been emphasized including contribution from our own group with regard to the in silico structure determination and structure-based small molecular drug development to treat cardiovascular disorders. The mainlining of neurotransmitter concentration using DBH as a drug target has placed dopamine and norepinephrine at the center of purpose, and these neurotransmitters deserve greater pharmacological attention. These catecholamines have also made anti-DBH drug discovery advantageous over other targets since it involves the physiological mechanisms for both BP and neural regulation. While clinical trials of nepicastat and disulfiram are already accomplished for the prevention of cocaine addiction via inhibition of neural DBH, ongoing or completed clinical trials for etamicastat and zamicastat for the prevention of hypertension via inhibition of peripheral DBH have established the importance of this unique drug target and provides glimpses of its promising future.

References

Abe M, Wu Z, Yamamoto M, Jin JJ, Tabara Y, Mogi M, Kohara K, Miki T, Nakura J (2005) Association of dopamine β-hydroxylase polymorphism with hypertension through interaction with fasting plasma glucose in Japanese. Hypertens Res 28(3):215–221

Almeida L, Nunes T, Costa R, Rocha JF, Vaz-da-Silva M, Soares-da-Silva P (2013) Etamicastat, a novel dopamine β-hydroxylase inhibitor: tolerability, pharmacokinetics, and pharmacodynamics in patients with hypertension. Clin Ther 35(12):1983–1996

Arnold AC, Garland EM, Celedonio JE, Raj SR, Abumrad NN, Biaggioni I, Robertson D, Luther JM, Shibao CA (2017) Hyperinsulinemia and insulin resistance in dopamine β-hydroxylase deficiency. J Clin Endocrinol Metab 102(1):10–14

Barrie ES, Weinshenker D, Verma A, Pendergrass SA, Lange LA, Ritchie MD, Wilson JG, Kuivaniemi H, Tromp G, Carey DJ (2014) Regulatory polymorphisms in human DBH affect peripheral gene expression and sympathetic activity. Circ Res 115(12):1017–1025

Beliaev A, Ferreira H, Learmonth DA, Soares-da-Silva P (2009) Dopamine β-monooxygenase: mechanism, substrates and inhibitors. Curr Enzym Inhib 5(1):27–43

Biaggioni I, Goldstein DS, Atkinson T, Robertson D (1990) Dopamine-β-hydroxylase deficiency in humans. Neurology 40(2):370–370

Bicker J, Alves G, Fortuna A, Soares-da-Silva P, Falcão A (2016) A new PAMPA model using an in-house brain lipid extract for screening the blood–brain barrier permeability of drug candidates. Int J Pharm 501(1-2):102–111

Bonifácio MJ, Sousa F, Neves M, Palma N, Igreja B, Pires NM, Wright LC, Soares-da-Silva P (2015) Characterization of the interaction of the novel antihypertensive etamicastat with human dopamine-β-hydroxylase: comparison with nepicastat. Eur J Pharmacol 751:50–58

Byrne CJ, Khurana S, Kumar A, Tai T (2018) Inflammatory signaling in hypertension: regulation of adrenal catecholamine biosynthesis. Front Endocrinol 9:343

Catelas DN, Serrão MP, Soares-Da-Silva P (2020) Effects of nepicastat upon dopamine-β-hydroxylase activity and dopamine and norepinephrine levels in the rat left ventricle, kidney, and adrenal gland. Clin Exp Hypertens 42(2):118–125

Chen Y, Wen G, Rao F, Zhang K, Wang L, Rodriguez-Flores JL, Sanchez AP, Mahata M, Taupenot L, Sun P (2010) Human dopamine beta-hydroxylase (DBH) regulatory polymorphism that influences enzymatic activity, autonomic function, and blood pressure. J Hypertens 28 (1):76–86

Claxton IM, Palfreyman MG, Poyser RH, Whiting RL (1976) BRL 8242 (2-[2-benzimidazolyl]-amino-2-imidazoline dihydrochloride), a new inhibitor of dopamine-β-hydroxylase with antihypertensive activity. Eur J Pharmacol 37(1):179–188

Corrodi H, Fuxe K, Hamberger B, Ljungdahl Å (1970) Studies on central and peripheral noradrenaline neurons using a new dopamine-β-hydroxylase inhibitor. Eur J Pharmacol 12(2):145–155

Cubells J, Kranzler H, McCance-Katz E, Anderson G, Malison R, Price L, Gelernter J (2000) A haplotype at the DBH locus, associated with low plasma dopamine β-hydroxylase activity, also associates with cocaine-induced paranoia. Mol Psychiatry 5(1):56–63

De La Garza IIR, Bubar MJ, Carbone CL, Moeller FG, Newton TF, Anastasio NC, Harper TA, Ware DL, Fuller MA, Holstein GJ (2015) Evaluation of the dopamine β-hydroxylase (DβH) inhibitor nepicastat in participants who meet criteria for cocaine use disorder. Prog Neuro-Psychopharmacol Biol Psychiatry 59:40–48

De Potter W, De Schaepdryver A, Smith A (1970) Release of chromogranin A and dopamine-beta-hydroxylase from adrenergic nerves during nerve stimulation. Acta Physiol Scand 357:8

De RC, Ranieri G, Bonfantino V, Adriani A, Filitti V, Ferrieri A (1993) Slow-release nicardipine in the treatment of arterial hypertension: comparative study vs. an ACE inhibitor. Minerva Cardioangiol 41(10):457–463

Deinum J, Steenbergen-Spanjers G, Jansen M, Boomsma F, Lenders J, Van Ittersum F, Hück N, van den Heuvel L, Wevers R (2004) DBH gene variants that cause low plasma dopamine β hydroxylase with or without a severe orthostatic syndrome. J Med Genet 41(4):e38

Del-Bel E, De-Miguel FF (2018) Extrasynaptic neurotransmission mediated by exocytosis and diffusive release of transmitter substances. Front Synaptic Neurosci 10:13

Dey SK, Saini M, Prabhakar P, Kundu S (2020) Dopamine β hydroxylase as a potential drug target to combat hypertension. Expert Opin Inv Drug (In Press)

Frigon RP, Stone RA (1978) Human plasma dopamine beta-hydroxylase. Purification and properties. J Biol Chem 253(19):6780–6786

Furuta Y, Washizaki M (1976) Effects of fusaric acid and its derivative on the cardiovascular system. Nihon Yakurigaku Zasshi 72(2):139–144

Garland EM (2012) Dopamine β-hydroxylase deficiency. In: Primer on the autonomic nervous system. Elsevier, Amsterdam, pp 431–434

Goldin L, Gershon E, Lake C, Murphy D, McGinniss M, Sparkes R (1982) Segregation and linkage studies of plasma dopamine-beta-hydroxylase (DBH), erythrocyte catechol-O-methyltransferase (COMT), and platelet monoamine oxidase (MAO): possible linkage between the ABO locus and a gene controlling DBH activity. Am J Hum Genet 34(2):250–262

Hachisu M, Tsuruoka T, Takahashi H, Asaoka H, Sekizawa Y, Koeda T, Inouye S (1983) Synthesis and antihypertensive activity of 5-O-substituted derivatives of 5-hydroxypicolinic acid. J Pharmacobiodyn 6(12):922–931

Hamblin KA, Flick-Smith H, Barnes KB, Pereira-Leal JB, Surkont J, Hampson R, Atkins HS, Harding SV (2019) Disulfiram, an alcohol dependence therapy, can inhibit the in vitro growth of Francisella tularensis. Int J Antimicrob Agents 54(1):85–88

Hammerschmidt T, Kummer MP, Terwel D, Martinez A, Gorji A, Pape H-C, Rommelfanger KS, Schroeder JP, Stoll M, Schultze J (2013) Selective loss of noradrenaline exacerbates early cognitive dysfunction and synaptic deficits in APP/PS1 mice. Biol Psychiatry 73(5):454–463

Hastings JA, Morris MJ, Lambert G, Lambert E, Esler M (2004) NPY and NPY Y1 receptor effects on noradrenaline overflow from the rat brain in vitro. Regul Pept 120(1-3):107–112

Houhou L, Lamouroux A, Biguet NF, Mallet J (1995) Expression of human dopamine β-hydroxylase in mammalian cells infected by recombinant vaccinia virus. Mechanisms for membrane attachment. J Biol Chem 270(21):12601–12606

Igreja B, Wright LC, Soares-da-Silva P (2016) Sustained high blood pressure reduction with etamicastat, a peripheral selective dopamine β-hydroxylase inhibitor. J Am Soc Hypertens 10 (3):207–216

Igreja B, Pires N, Loureiro A, Wright L, Soares-da-Silva P (2019a) Cardiometabolic and inflammatory benefits of sympathetic down-regulation with zamicastat in aged spontaneously hypertensive rats. ACS Pharmacol Trans Sci 2(5):353–360

Igreja B, Pires NM, Wright LC, Soares-da-Silva P (2019b) Effects of zamicastat treatment in a genetic model of salt-sensitive hypertension and heart failure. Eur J Pharmacol 842:125–132

Ishii Y, Fujii Y, Mimura C, Umezawa H (1975) Pharmacological action of FD-008, a new dopamine beta-hydroxylase inhibitor. I. Effects on blood pressure in rats and dogs. Arzneimittelforschung 25(1):55–59

Jepma M, Deinum J, Asplund CL, Rombouts SA, Tamsma JT, Tjeerdema N, Spapé MM, Garland EM, Robertson D, Lenders JW (2011) Neurocognitive function in dopamine-β-hydroxylase deficiency. Neuropsychopharmacology 36(8):1608–1619

Johnson G, Boukma S, Kim E (1970) In vivo inhibition of dopamine β-hydroxylase by 1-phenyl-3-(2-thiazolyl)-2-thiourea (U-14,624). J Pharmacol Exp Ther 171(1):80–87

Johnson DG, Thoa NB, Weinshilboum R, Axelrod J, Kopin IJ (1971) Enhanced release of dopamine-β-hydroxylase from sympathetic nerves by calcium and phenoxybenzamine and its reversal by prostaglandins. Proc Natl Acad Sci U S A 68(9):2227–2230

Kapoor A, Shandilya M, Kundu S (2011) Structural insight of dopamine β-hydroxylase, a drug target for complex traits, and functional significance of exonic single nucleotide polymorphisms. PLoS One 6(10):e26509

Kemper C, O'connor D, Westlund K (1987) Immunocytochemical localization of dopamine-β-hydroxylase in neurons of the human brain stem. Neuroscience 23(3):981–989

Kim C-H, Leung A, Huh YH, Yang E, Kim D-J, Leblanc P, Ryu H, Kim K, Kim D-W, Garland EM (2011) Norepinephrine deficiency is caused by combined abnormal mRNA processing and defective protein trafficking of dopamine β-hydroxylase. J Biol Chem 286(11):9196–9204

Kiriakov A, Khlebarova M, Staneva-Stoicheva D, Panova I (1973) The effect of prolonged treatment hypertensive rats with antihypertensive agents with different mechanisms of action on blood pressure and noradrenaline concentration in the myocardium, brain and aorta. Eksp Med Morfol 12(3):135–141

Kobayashi K, Kurosawa Y, Fujita K, Nagatsu T (1989) Human dopamine β-hydroxylase gene: two mRNA types having different 3'terminal regions are produced through alternative polyadenylation. Nucleic Acids Res 17(3):1089–1102

Kosten TR, Wu G, Huang W, Harding MJ, Hamon SC, Lappalainen J, Nielsen DA (2013) Pharmacogenetic randomized trial for cocaine abuse: disulfiram and dopamine β-hydroxylase. Biol Psychiatry 73(3):219–224

Kruse LI, Kaiser C, DeWolf WE Jr, Frazee JS, Erickson RW, Ezekiel M, Ohlstein EH, Ruffolo RR Jr, Berkowitz BA (1986) Substituted 1-benzylimidazole-2-thiols as potent and orally active inhibitors of dopamine. Beta-hydroxylase. J Med Chem 29(6):887–889

Kruse LI, Kaiser C, DeWolf WE Jr, Frazee JS, Ross ST, Wawro J, Wise M, Flaim KE, Sawyer JL (1987) Multisubstrate inhibitors of dopamine. Beta-hydroxylase. 2. Structure-activity relationships at the phenethylamine binding site. J Med Chem 30(3):486–494

Kundu S, Thelma B, Maulik S, Prabhakar P, Dey SK (2017) Novel anti-hypertensive and anti-cardiac hypertrophic compounds. Indian Patent Application. 201711036983A

Kundu S, Dey SK, Thelma B, Kovuru G, Prabhakar P, Saini M (2018) An anti-hypertensive cardio-protective composition. Indian Patent Application. 201811005899A

Laduron PM (1975) Evidence for a localization of dopamine-β-hydroxylase within the chromaffin granules. FEBS Lett 52(1):132–134

Lerner P, Goodwin F, Post R, Major L, Ballenger J, Lovenberg W (1978) Dopamine-beta-hydroxylase in the cerebrospinal fluid of psychiatric patients. Biol Psychiatry 13(6):685–694

Levin M (1961) The levels of the nervous system and their capacity to function idependently of each other. J Nerv Ment Dis 132(1):75–79

Lewis EJ, Asnani L (1992) Soluble and membrane-bound forms of dopamine beta-hydroxylase are encoded by the same mRNA. J Biol Chem 267(1):494–500

Lippmann W, Lloyd K (1969) Dopamine-ß-hydroxylase inhibition by dimethyldithiocarbamate and related compounds. Biochem Pharmacol 18(10):2507–2516

Ljones T, Flatmark T (1974) Dophamine β-Hhydroxylase: evidence against a ping-pong mechanism. FEBS Lett 49(1):49–52

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) Molecular cell biology, 4th edn. W. H. Freeman and Company, New York

Loureiro A, Soares-da-Silva P (2015) Distribution and pharmacokinetics of etamicastat and its N-acetylated metabolite (BIA 5-961) in dog and monkey. Xenobiotica 45(10):903–911

Loureiro AI, Bonifácio MJ, Fernandes-Lopes C, Pires N, Igreja B, Wright LC, Soares-da-Silva P (2015) Role of P-glycoprotein and permeability upon the brain distribution and pharmacodynamics of etamicastat: a comparison with nepicastat. Xenobiotica 45(9):828–839

Marino MD, Bourdélat-Parks BN, Liles LC, Weinshenker D (2005) Genetic reduction of noradrenergic function alters social memory and reduces aggression in mice. Behav Brain Res 161 (2):197–203

Mathias C, Bannister R, Cortelli P, Heslop K, Polak J, Raimbach S, Springall D, Watson L (1990) Clinical, autonomic and therapeutic observations in two siblings with postural hypotension and sympathetic failure due to an inability to synthesize noradrenaline from dopamine because of a deficiency of dopamine beta hydroxylase. Q J Med 75(3):617–633

Matsuzaki M, Nakamura K, Akutsu S, Onodera K, Sekino M (1976) Fundamental studies on fusaric acid and calcium fusarate. Acute toxicity and antihypertensive effects. Jpn J Antibiot 29 (5):439–455

Matuzas W, Meltzer H, Uhlenhuth E, Glass R, Tong C (1982) Plasma dopamine-beta-hydroxylase in depressed patients. Biol Psychiatry 17(12):1415–1424

McCorry LK (2007) Physiology of the autonomic nervous system. Am J Pharm Educ 71(4):78

Molinoff PB, Landsberg L, Axelrod J (1969) An enzymatic assay for octopamine and other β-hydroxylated phenylethylamines. J Pharmacol Exp Ther 170(2):253–261

Nagatsu T (2009) Simple photometric assay of dopamine-β-hydroxylase activity in human blood: useful in clinical chemistry. Clin Chem 55(1):193–194

Nagatsu T, Kuzuya H, Hidaka H (1967) Inhibition of dopamine β-hydroxylase by sulfhydryl compounds and the nature of the natural inhibitors. Biochim Biophys Acta 139(2):319–327

Nunes T, Rocha JF, Vaz-da-Silva M, Igreja B, Wright LC, Falcão A, Almeida L, Soares-da-Silva P (2010) Safety, tolerability, and pharmacokinetics of etamicastat, a novel dopamine-β-hydroxylase inhibitor, in a rising multiple-dose study in young healthy subjects. Drugs R D 10(4):225–242

Ohlstein E, Kruse L, Ezekiel M, Sherman S, Erickson R, DeWolf W, Berkowitz B (1987) Cardiovascular effects of a new potent dopamine beta-hydroxylase inhibitor in spontaneously hypertensive rats. J Pharmacol Exp Ther 241(2):554–559

Parasuraman R, Greenwood PM, Kumar R, Fossella J (2005) Beyond heritability: neurotransmitter genes differentially modulate visuospatial attention and working memory. Psychol Sci 16 (3):200–207

Parasuraman R, de Visser E, Lin M-K, Greenwood PM (2012) Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. PLoS One 7(6):e39675

Pires NM, Igreja B, Moura E, Wright LC, Serrão MP, Soares-da-Silva P (2015) Blood pressure decrease in spontaneously hypertensive rats folowing renal denervation or dopamine β-hydroxylase inhibition with etamicastat. Hypertens Res 38(9):605–612

Punchaichira TJ, Dey SK, Mukhopadhyay A, Kundu S, Thelma B (2017) Characterization of SNPs in the dopamine-β-hydroxylase gene providing new insights into its structure-function relationship. Neurogenetics 18(3):155–168

Punchaichira TJ, Deshpande SN, Thelma B (2018) Determination of dopamine-β-hydroxylase activity in human serum using UHPLC-PDA detection. Neurochem Res 43(12):2324–2332

Rapacciuolo A, Esposito G, Caron K, Mao L, Thomas SA, Rockman HA (2001) Important role of endogenous norepinephrine and epinephrine in the development of in vivo pressure-overload cardiac hypertrophy. J Am Coll Cardiol 38(3):876–882

Robbins TW, Arnsten AF (2009) The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. Annu Rev Neurosci 32:267–287

Robertson D, Goldberg MR, Onrot J, Hollister AS, Wiley R, Thompson JG Jr, Robertson RM (1986) Isolated failure of autonomic noradrenergic neurotransmission. N Engl J Med 314 (23):1494–1497

Rocha JF, Vaz-Da-Silva M, Nunes T, Igreja B, Loureiro AI, Bonifácio MJ, Wright LC, Falcão A, Almeida L, Soares-Da-Silva P (2012) Single-dose tolerability, pharmacokinetics, and pharmacodynamics of etamicastat (BIA 5–453), a new dopamine β-hydroxylase inhibitor, in healthy subjects. J Clin Pharmacol 52(2):156–170

Rosenberg RC, Lovenberg W (1977) Active dimers of dopamine β-hydroxylase in human plasma. Mol Pharmacol 13(4):652–661

Ross MO, Rosenzweig AC (2017) A tale of two methane monooxygenases. J Biol Inorg Chem 22 (2-3):307–319

Rush R, Geffen L (1980) Dopamine β-hydroxylase in health and disease. Crit Rev Clin Lab Sci 12 (3):241–277

Sabbah HN, Stanley WC, Sharov VG, Mishima T, Tanimura M, Benedict CR, Hegde S, Goldstein S (2000) Effects of dopamine β-hydroxylase inhibition with nepicastat on the progression of left ventricular dysfunction and remodeling in dogs with chronic heart failure. Circulation 102 (16):1990–1995

Saxena A, Fleming PJ (1983) Isolation and reconstitution of the membrane-bound form of dopamine beta-hydroxylase. J Biol Chem 258(7):4147–4152

Schroeder JP, Cooper DA, Schank JR, Lyle MA, Gaval-Cruz M, Ogbonmwan YE, Pozdeyev N, Freeman KG, Iuvone PM, Edwards GL (2010) Disulfiram attenuates drug-primed reinstatement of cocaine seeking via inhibition of dopamine β-hydroxylase. Neuropsychopharmacology 35 (12):2440–2449

Schroeder JP, Epps SA, Grice TW, Weinshenker D (2013) The selective dopamine β-hydroxylase inhibitor nepicastat attenuates multiple aspects of cocaine-seeking behavior. Neuropsychopharmacology 38(6):1032–1038

Seo H, Yang C, Kim H-S, Kim K-S (1996) Multiple protein factors interact with the cis-regulatory elements of the proximal promoter in a cell-specific manner and regulate transcription of the dopamine b-hydroxylase gene. J Neurosci 16(13):4102–4112

Shen B, Cheng K-T, Leung Y-K, Kwok Y-C, Kwan H-Y, Wong C-O, Chen Z-Y, Huang Y, Yao X (2008) Epinephrine-induced Ca2+ influx in vascular endothelial cells is mediated by CNGA2 channels. J Mol Cell Cardiol 45(3):437–445

Slater EP, Zaremba S, Hogue-Angeletti RA (1981) Purification of membrane-bound dopamine β-monooxygenase from chromaffin granules: relation to soluble dopamine β-monooxygenase. Arch Biochem Biophys 211(1):288–296

Sokoloff RL, Frigon RP, O'Connor DT (1985) Dopamine-β-hydroxylase: structural comparisons of membrane-bound versus soluble forms from adrenal medulla and pheochromocytoma. J Neurochem 44(2):411–420

Stewart MH, Lavie CJ, Ventura HO (2018) Future pharmacological therapy in hypertension. Curr Opin Cardiol 33(4):408–415

Stubbusch J, Majdazari A, Schmidt M, Schütz G, Deller T, Rohrer H (2011) Generation of the tamoxifen-inducible DBH-Cre transgenic mouse line DBH-CT. Genesis 49(12):935–941

Tang S, Yao B, Li N, Lin S, Huang Z (2018) Association of dopamine beta-hydroxylase polymorphisms with alzheimer's disease, Parkinson's disease and schizophrenia: evidence based on currently available loci. Cell Physiol Biochem 51(1):411–428

Taylor CS, Fleming PJ (1989) Conversion of soluble dopamine beta-hydroxylase to a membrane binding form. J Biol Chem 264(26):15242–15246

Teitelman G, Baker H, Joh TH, Reis DJ (1979) Appearance of catecholamine-synthesizing enzymes during development of rat sympathetic nervous system: possible role of tissue environment. Proc Natl Acad Sci U S A 76(1):509–513

Tishchenko K, Beloglazkina E, Mazhuga A, Zyk N (2016) Copper-containing enzymes: site types and low-molecular-weight model compounds. Rev J Chem 6(1):49–82

Veld AM, Moleman P, Boomsma F, Schalekamp M (1987) Congenital dopamine-beta-hydroxylase deficiency: a novel orthostatic syndrome. Lancet 329(8526):183–188

Vendelboe TV, Harris P, Zhao Y, Walter TS, Harlos K, El Omari K, Christensen HE (2016) The crystal structure of human dopamine β-hydroxylase at 2.9 Å resolution. Sci Adv 2(4):e1500980

Wallace EF, Krantz MJ, Lovenberg W (1973) Dopamine-β-hydroxylase: a tetrameric glycoprotein. Proc Natl Acad Sci U S A 70(8):2253–2255

Wang H, Ng T (1999) Pharmacological activities of fusaric acid (5-butylpicolinic acid). Life Sci 65 (9):849–856

Weinshilboum RM (1989) Catecholamine biochemical genetics. In: Trendelenburg U, Weiner N (eds) Catecholamines II, pp 391–425. Handbook of experimental pharmacology, vol 90/2. Springer, Berlin, Heidelberg

Weinshilboum RM, Thoa NB, Johnson DG, Kopin IJ, Axelrod J (1971) Proportional release of norepinephrine and dopamine-β-hydroxylase from sympathetic nerves. Science 174 (4016):1349–1351

Whaley-Connell A, Sowers K, Sowers JR (2006) Hypertension and cardiovascular disease. In: The diabetic kidney. Springer, Berlin, pp 499–513

William Tank A, Lee Wong D (2011) Peripheral and central effects of circulating catecholamines. Compr Physiol 5(1):1–15

Zabetian CP, Anderson GM, Buxbaum SG, Elston RC, Ichinose H, Nagatsu T, Kim K-S, Kim C-H, Malison RT, Gelernter J (2001) A quantitative-trait analysis of human plasma–dopamine β-hydroxylase activity: evidence for a major functional polymorphism at the DBH locus. Am J Hum Genet 68(2):515–522

Molecular Motors: Subdomain Dynamics and Mechanochemistry

15

Meenakshi Singh and Sudhir Kumar Singh

Abstract

Biological cells contain nano-molecular motors that perform essential functions such as intracellular transport, muscle contraction, and chromosome separation. Molecular motors are enzymatic proteins that drive the intracellular trafficking by converting the chemical energy of adenosine triphosphate (ATP) hydrolysis into mechanical action. Among different motor proteins coexisting in every eukaryotic cell, cytoplasmic motor proteins are plausibly most fascinating. These proteins bind to a polarized cytoskeleton filament, move unidirectionally and divided into three motor classes: myosins, which move on actin filaments, and dyneins and kinesins, which use microtubules (MT) as tracks. ATP hydrolysis alters their subdomain dynamics in the catalytic domain which is further communicated to the track-binding site. This chapter will focus on kinesins, the structural and molecular basis of force generation, how they differ markedly from myosins and dyneins and, insights into their remarkable motor mechanochemistry. We will discuss the core architecture and structural elements of kinesin and how the intramolecular communication within the kinesin motor domain translates into a large conformational change that leads to a directional movement along the microtubule track. Concomitantly, recent findings of the bidirectional motility of kinesin-5 motors will also be discussed in detail that is contrary to the previous dogma of unidirectional movement of plus end-directed molecular motors. We will also substantiate several structural determinants of kinesin-5 motors that regulate directional switching along with the evidence of cover-neck bundle formation in yeast kinesin-5 that is well established for the plus end-

M. Singh
Department of Medicinal Chemistry, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India

S. K. Singh (✉)
Department of Chemistry, The Marcus Family Campus, Ben-Gurion University of the Negev, Beersheba, Israel

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_15

directed movement in kinesin-1 motors. Further, we brief about some small molecule inhibitors that bind to Loop5 of kinesin-5 and affect the subdomain dynamics and important for anti-cancer treatment. The understanding of structural and biophysical dynamics of kinesin motors could be helpful to elucidate how these motors function during mitosis and the molecular mechanism of bidirectionality and force generation.

Keywords

Kinesins · Microtubules · Motor domain · Kinesin-1 · Kinesin-5 · Mitosis · Bidirectionality · Neck linker

15.1 Molecular Motor: a Brief Introduction

Biological cells are sustained by the interplay of several cellular functions, which are extremely complex and dynamic in nature. Some of these essential processes are gene transcription, translation, cell division, intracellular transport, muscle contraction, etc. The biophysical and biochemical basis of these various processes is still not fully understood. Several proteins involved in these processes are active enzymatic molecules, generally called motor proteins or molecular motors. These protein-based motors transform the chemical energy of adenosine triphosphate (ATP) into mechanical work. Understanding nanometer-sized molecular machines mode of action would provide insight into fundamental cellular processes and further help in designing a strategy to combat some of the deadly diseases that occur due to malfunction of the cellular machinery.

15.1.1 Classifications of Motor Proteins

There are various enzymatic motor proteins which are grouped based upon their function and binding to substrates which are dynamic mostly but in some cases, static also. These substrates are cellular structures consisting of cytoskeleton tubulin filaments, nucleic acids, membranes, etc.

15.1.1.1 Cytoskeleton Filaments Motor Proteins

This includes molecular motors such as dyneins, myosins, and kinesins. They bind and move along the cytoskeleton filaments (actin filaments and microtubules) utilizing the energy of hydrolysis of ATP. Actin motors such as myosin move along microfilaments through interaction with actin, and microtubule motors such as dynein and kinesin move along microtubules (MTs) through interaction with tubulin. They are the leading players in cellular transport processes (Howard 2002; Roberts et al. 2013; Thompson and Langford 2002; Endow et al. 2010).

15.1.1.2 Nucleic Acids Motor Proteins

This includes DNA and RNA polymerases, topoisomerases, gyrases, helicases, etc. These motor proteins bind to DNA and RNA molecules, using ATP as a source of chemical energy. They are involved in transcription and translational machinery for dispensing the genetic information (Sutton and Walker 2001; Parker 2000; Levine et al. 1998).

15.1.1.3 Rotary Motor Proteins

These include the bacterial flagellar rotary motor (essential for bacterial migration and motility) and F0F1-ATP synthase, which synthesizes ATP molecules in mitochondria and provide energy in the living system. They are usually bound to cellular membranes (Berry and Armitage 1999; Minamino et al. 2008; Sowa et al. 2005; Jonckheere et al. 2012; Neupane et al. 2019).

15.1.2 Kinesins, Myosins, and Dyneins

All the three families of motor proteins convert the chemical energy of ATP to the kinetic energy of movement and subsequently power most eukaryotic cellular movements. Myosin and kinesin are Ras family guanosine triphosphatases (GTPases) and seems to have shared a common ancestor during evolution, whereas dynein belongs to AAA adenosine triphosphatase (ATPase) family. Myosin walks on actin filaments, whereas kinesin and dynein use MTs and move toward the plus and minus ends, respectively (Kull and Endow 2013; Roberts et al. 2013). The core architecture of the motor domains of the proteins myosin and kinesin is structurally similar, and they have similar force-producing mechanisms that are generated due to the conformational change in their motor domain by nucleotide hydrolysis, i.e., ATP binding, hydrolysis, Pi release and further ADP release (Kull and Endow 2013). It is quite amazing that despite no sequence identity, they share a similar mechanism of force generation. The motor mechanochemical cycle of kinesin will be discussed in a later section.

Kinesins and dyneins, are both involved in moving cargo along MTs, but an important difference is that most kinesins travel toward the plus end of the MT, i.e., away from the center of the cell, while dyneins travel toward the minus end of the microtubule, i.e., toward the center of the cell. Thus, kinesins function to bring cargoes to the periphery of the cell, while dyneins function to carry cargoes to the center of the cell. Contrary to myosin and kinesin, the dynein shows dissimilarity at the structural level (Burgess et al. 2003) and subsequently at the level of force generation as well (Roberts et al. 2009). Dyneins consist of two groups, flagellar and cytoplasmic dyneins where flagellar dyneins root force to beat flagella/cilia, while cytoplasmic dyneins are involved in cargo transport within a cell.

15.1.3 Kinesin Preview

Kinesin was discovered about 38 years ago in 1985 (Allen et al. 1982; Brady et al. 1982; Vale et al. 1985), which is now known as kinesin-1 and considered as the founding member of the conventional kinesin. Kinesin-1 is a heterotetrameric protein consisting of two kinesin heavy chains (KHCs) and two kinesin light chains (KLCs) (Fig. 15.1). The N-terminal region of approximately 350 amino acids has a motor activity governed by ATP binding which leads to conformational rearrangements and further this ATPase-dependent movement helps the kinesin to move upon the MTs. The motor domain is nearly conserved among the kinesin-related motor proteins. The motor domain is followed by the KHC dimerization domain, which contains the alpha-helical coiled-coil region. Further, toward the C-terminal, the KLCs tail domain or cargo binding domain of kinesin-1 is present. KLC has a region of six tetratricopeptide repeat (TPR) units, which are also conserved and predicted to form a triple alpha-helical groove. The study showed that kinesin-1 has an essential role in vesicle transport in neuronal cells (Saxton et al. 1991). Afterward, several homologs of kinesin-1 have been found in a wide range of organisms, i.e., from yeast, *Caenorhabditis elegans*, Drosophila, mouse, and human.

In the early 1990s, many other kinesin-related proteins were reported. Consequently, a standardized system of kinesin nomenclature was introduced in 2004, renaming the different kinesins groups by number (Lawrence et al. 2004). There are currently 14 recognized kinesin families, and some of them are shown in Fig. 15.2. Kinesin motor proteins have mainly more diverse tails domain, allowing for specific cellular functions of each kinesin motor isoform.

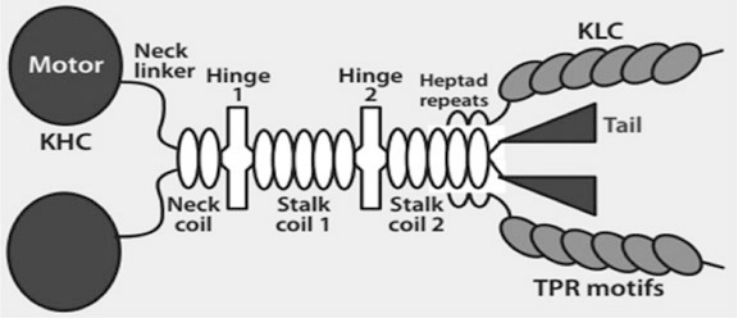


Fig. 15.1 Schematic organization of kinesin-1 structure. Subunit composition of the heterotetrameric kinesin-1 motor contains a motor domain (*dark green oval*) at their N-terminus for ATP binding site and the processive motion toward the plus ends of MTs. This region is considered as KHC. These kinesins also have a neck domain (neck linker and neck coil) and varying amounts of coiled-coil stalk regions for oligomerization and followed by KLC. Both KHC and KLC may be involved in cargo binding either together or individually. They also have protein–protein or protein–lipid interaction domains tetratricopeptide repeat (TPR) that participate in cargo binding. The figure is adapted from Verhey et al. (2011)

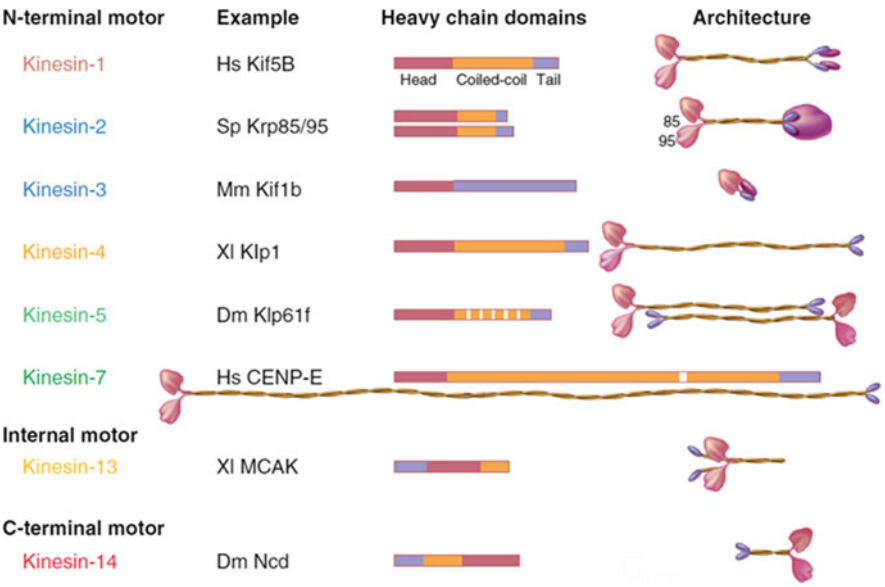


Fig. 15.2 Schematic structure of different kinesin families. It is based on the phylogenic analysis, sequence of motor domain, domain structure of kinesin heavy chain and light chain domains (Kim and Endow 2000). The schematic models of different kinesins have different architecture, and they contain catalytic domain (red), coiled-coil tail (orange), and tail piece (blue). Figure adapted from Corden et al. (2017)

15.1.4 Function of Kinesins

Kinesins transport a variety of cargo, including organelles and chromosomes, along MTs. The role and function of different kinesins are as follows.

15.1.4.1 Kinesin-1

(Representative members—LpKHC, DmKHC, KIF5B, KHC, NKin, KLP1, KinA, and DdK5): Kinesin-1 members involve in anterograde movement and play important roles in the transport of synaptic vesicles and the maintenance of neuronal architecture (Corden et al. 2017; Saxton et al. 1991).

15.1.4.2 Kinesin-2

(Representative members—StrPuKRP85/95, KIF3A/3B, KIF17, Krp85/95, Osm3, Fla10): It is involved in a wide variety of transport events such as the movement of particles toward the tip of the axoneme in cilia and flagella, their maintenance, to transport along cytoplasmic MTs of organelles, membrane-bound vesicles, mRNA granules, and melanosomes (Scholey 2013, 2008; Kozminski et al. 1993; Hirokawa et al. 2009).

15.1.4.3 Kinesin-3

(Representative members—MmKIF1B, KIF1A, KIF1B, KIF13A, KIF16B, UNC104): The family members are fast organelle transporters in the amoeba (Pollock et al. 1999) and endosome transporters in fungi (Seidel et al. 2013). Kinesin-3 members also act as transporters of viral particles, vesicles and mitochondria in mammalian cells as well as involved in the orderly progression of cell division (Lo et al. 2011; Nangaku et al. 1994; Kratchmarov et al. 2013).

15.1.4.4 Kinesin-4

(Representative members—MmKIF4A, KIF4A, KIF21A/B, chromokinesin): Kinesin-4 motors bind both DNA and MTs. Kinesin-4 is involved in the regulation of chromatin and chromosome structure, replication, and DNA repair (Mazumdar et al. 2011). They are also involved in the regulation and stability of spindle length and hence regulate mitosis and meiosis (Nunes Bastos et al. 2013; Hu et al. 2011; Heath and Wignall 2019).

15.1.4.5 Kinesin-5

(Representative members—KIF11, Eg5, BimC, CIN8, KIP1, Cut7): Bipolar Kinesin-5 play essential roles in spindle assembly and function by generating outward forces and phosphor-regulation which establish and maintain spindle bipolarity and contribute to microtubule flux (Mann and Wadsworth 2019; Singh et al. 2018).

15.1.4.6 Kinesin-6

(Representative members—KIF20, KIF23, Rab6Kinesin, CHO1, MKLP1, Zen4, MPP1): Kinesin-6 family members are localized to the spindle midzone during anaphase B, and they contribute to anaphase B spindle elongation and cytokinesis (Adams et al. 1998; Hardin 2012; Janisch et al. 2018).

15.1.4.7 Kinesin-7

(Representative members—KIF10, CENP-E, CMET, CANA, KIP2): The motor proteins congregate at kinetochores and assists the movement of the chromosome toward the middle of the mitotic spindle prometaphase (Yardimci et al. 2008; Yu et al. 2019).

15.1.4.8 Kinesin-8

(Representative members—KIF18A/18B KIF19A, KLP67A, KIP3): Kinesin 8 family has been shown to play an important role in chromosome alignment during mitosis (Mayr et al. 2007; Stumpff et al. 2008). They control the MT length in a variety of cellular processes by removing tubulin dimers from the ends of MTs as they are also termed as plus-end MT-destabilizing enzymes (Varga et al. 2006).

15.1.4.9 Kinesin-9

(Representative members—KIF6, KIF9, KRP3, CrKLP1): Kinesin-9 is strongly associated with the flagellar skeleton and participate in flagellar motility in

Trypanosoma brucei (Demonchy et al. 2009), and they are also engaged during spermatid differentiation and ciliogenesis in semi-aquatic fern, *Marsileavestita* (Tomei and Wolniak 2016).

15.1.4.10 Kinesin-10

(Representative members—KIF22, KID, Nod): The kinesin-10 family members are commonly referred to as "Kid" in humans and "KIF 22" in the mouse. They have been suggested to be involved in chromosomal movement along MTs during prometaphase and metaphase, maintenance of proper metaphase spindle size as well as the orientation of chromosome arms and chromosome oscillation during metaphase (Tokai-Nishizumi et al. 2005; Miki et al. 2005; Levesque and Compton 2001; Antonio et al. 2000).

15.1.4.11 Kinesin-11

(Representative members—KIF26A, KIF26B, VAB8, SMY1): Kinesin-11 family member Smy1 involves in the transport of secretory vesicles and functions closely with a myosin-V and its receptor in the transportation of a specific cargo and actin assembly in the yeast (Lwin et al. 2016). Another member vab-8 is involved in regulating and controlling the direction of cell and axon growth cone migrations in *C. elegans* (Wolf et al. 1998). These kinesins do not bind to MTs, and the catalytic core is highly divergent compared with those of other families of kinesin. The role of the catalytic core in this family has not been explored to date.

15.1.4.12 Kinesin-12

(Representative members—KIF12, KIF15, HKLP2, KLP54D, Xklp2, PAKRPd): Kinesin-12 is a mitotic MT-associated motor protein which affects axonal growth and branching in the rat (Liu et al. 2010). Like kinesin-5, they also play a crucial role in spindle assembly. In plants, kinesin-12 members are involved in critical events during cell division and mediate in several development processes such as male gametophyte, embryo, seedling, and seed development (Muller and Livanos 2019; Tian et al. 2016; Drechsler and McAinsh 2016).

15.1.4.13 Kinesin-13

(Representative members—KIF2A, MCAK, XKCM1, PfKinI): Kinesin-13 members can travel to both the minus and plus ends of MTs, whereas most motors are unidirectional. Kinesin-13 has a role in vesicle transport. They also drive sister chromatid separation during anaphase in drosophila (Rogers et al. 2004). They catalyze the depolymerization of MTs by utilizing the ATP hydrolysis to remove tubulin dimers from the ends of MTs (Ogawa et al. 2004; Wang et al. 2017; Moores et al. 2006).

15.1.4.14 Kinesin-14

(Representative members—KIFC1, CHO2, Ncd, Kar3, KatA): Kinesin-14 regulates the fundamental function of mitosis such as regulation of microtubule organization, spindle assembly, and chromosome segregation (Walczak et al. 1997). These are

minus-end-directed motors along MT. Due to the generation of inward pulling forces on spindles by kinesin-14 members, there is mutual antagonism between kinesin-14 and kinesin-5 family members during spindle assembly and MT nucleation (Yukawa et al. 2015; Hentrich and Surrey 2010).

15.1.5 Kinesin Directional Movement

Kinesin motor proteins travel in a specific direction along the MTs, which are tubule-like structures with a diameter of 25 nm (Snyder and McIntosh 1976), formed by α, β-tubulin heterodimers. These tubulin polymers form part of the cytoskeleton and provide shape and structure to the bacteria and cytoplasm of eukaryotic cells. The most common form of a microtubule consists of 13 protofilaments in the tubular arrangement that forms the microtubule wall, to which several MT-associated proteins and motor proteins bind. The motor domain heads of kinesins only bind to the MT in one orientation. At the same time, ATP binding gives each step its direction through a process known as neck linker zippering, which is NL docked conformation localized to the main body of the motor domain (Rice et al. 1999).

The kinesin motor can be divided based upon the location of its motor domain. It is termed as N-type if it is located on the N-terminus, C-terminus (C-type), or middle region (M-type). Kinesin-1 (KIF5B) and kinesin-5 (Eg5) are the most common examples of the N-type kinesins and move toward the plus end of MTs; the C-type kinesin consists of Drosophila Ncd motor that walks to the minus end; and the M-type kinesin, i.e., kinesin-13, travels to both ends and depolymerizes MTs (Hirokawa 1998). Kinesins move cargo toward the positive (+) end of an MT, also known as anterograde transport/orthograde transport. Contrary to this, kinesin-14 family proteins (such as *Drosophila melanogaster* Ncd, budding yeast Kar3, and *Arabidopsis thaliana* Atk5) walk in the opposite direction, i.e., toward the MT minus end (Ambrose et al. 2005). They transport cargo from the periphery of the cell toward the center termed as a retrograde movement. The dyneins also move toward the minus end of the MT. However, a recent study demonstrates that some kinesins are bidirectional in nature and are more divergent than previously thought: Yeast kinesin-5, Cin8, Kip1, and Cut7 are bidirectional motors (Thiede et al. 2012; Singh et al. 2018; Gerson-Gurwitz et al. 2011; Edamatsu 2014), and the kinesin-14 KlpA is a plus-end-directed motor on single MT (Popchock et al. 2017).

15.1.6 Kinesin Structural Elements: General Architecture

The structural study of kinesin and myosin elucidated the similarity of core architecture and nucleotide-binding site for the power stroke generation (Kull et al. 1996; Rayment et al. 1993). The kinesin motor domain contains the following: (1) the motor core consisting of an αβα-fold with three α-helices on each side of an eight-stranded β-sheet; (2) a family-specific neck-linker sequence immediately preceding

or following the motor core; (3) the cover strand (CS), a motif at the opposite terminus relative to the NL.

Based upon the several structural and biochemical studies, such as a combination of cryo-electron microscopy (Cryo-EM), X-ray crystallography, single molecule spectroscopy, pre-steady-state kinetics, and site-directed mutagenesis techniques, the critical mechanical elements in kinesins have been identified which undergo nucleotide- and microtubule-dependent conformational changes. These structural elements further allowed the scientific researchers to construct a structural and mechanistic model that could explain the mechanism of kinesin movement along the MTs.

15.1.6.1 Nucleotide-Dependent Structural Domain

The structural studies have revealed that a kinesin motor domain comprises of three subdomains (the P-loop, switch 1, and switch 2) that undergo conformational changes by coupling ATP hydrolysis and mediate directional movement along MTs. The ATP binding site in the kinesin motor domain consists of four motifs, common to P-loop containing proteins (Walker et al. 1982) that are also found in myosin and G-proteins. The P-loop is also called as Walker A motif or Walker loop, or phosphate-binding loop that is associated with phosphate binding. The Walker A motif has the sequence consensus motif (GxxxxGKT/S) and forms a phosphate-binding loop (P-loop) between the β3 strand of the central β-sheet and helix α2. The P-loop of kinesins has two extra conserved residues and the motif pattern is GQTxxGKS/T (Sack et al. 1999). This loop tightly binds the β-phosphate of the nucleotide. The other two motifs are, the switch-1 (NxxSSR) and the switch-2 motif (DxxGxE) that changes their conformation and interaction in response to the presence or absence of γ-phosphate. The P-loop interacts with Mg^{2+} ions and α-, β- and γ-phosphates of nucleotides and both switch I and switch II act as sensors that recognize the existence of γ-phosphate (Gigant et al. 2013). Hence, a small conformational change of few angstroms in the nucleotide-binding pocket is amplified into the large displacements of several nanometers associated with motility along the MTs. Figure 15.3a depicts the binding of nucleotides ATP and ADP to kinesin triggers conformational changes that also lead to a difference in the structure attained by Switch1/L9 loop, Switch2/L11 loop, and P-loop at a different stage.

15.1.6.2 MT-Dependent Structural Domain

1-Neck Linker

This structural element is present outside of the motor catalytic domain. The kinesin's NL locates between the α6-helix of the motor domain and the α7-helix of the coiled-coil. The NL (β9 and β10) is usually a ∼15–18 long amino acid segment present at the carboxy-terminal to the catalytic core that connects the motor domain to the N-terminal of the coiled-coil dimerization domain (Fig. 15.1). The sequence of NL is highly conserved among plus-end-directed motors and considered as one of the principal factors for kinesin force generation and energy transduction (Case et al. 1997). The two motor domains of dimeric kinesins

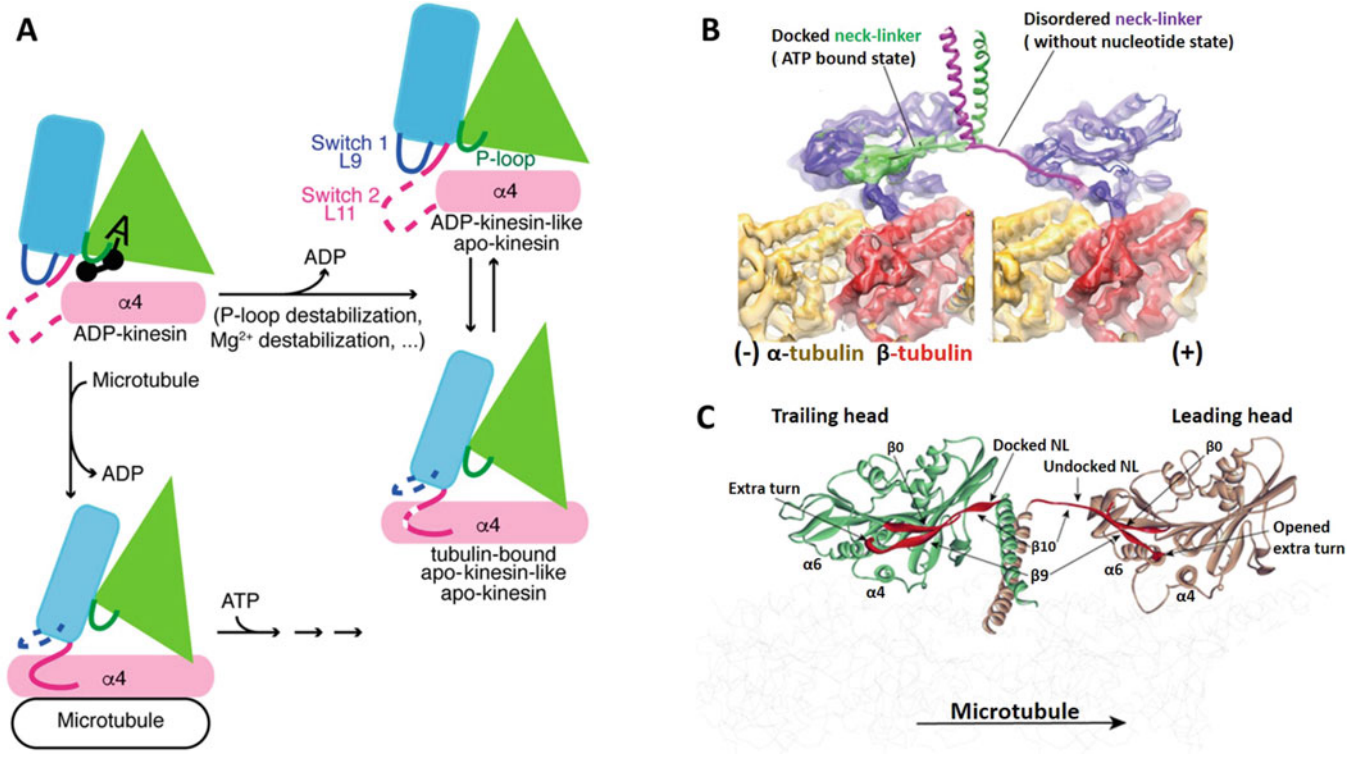


Fig. 15.3 Nucleotide and MT-dependent change in structural conformation. (**a**) The binding of nucleotide ATP or ADP to kinesin-1 and microtubules triggers conformational changes. Subdomain orientation changes during the kinesin mechanochemical cycle. Switch1/L9 loop (blue), Switch2/L11 loop (magenta), P-loop (green). Figure adapted from Cao et al. (2017). (**b**) Neck-linker conformational change. Cryo-EM three-dimensional (3-D) reconstructions show the kinesin-1 heads (blue) bound to an MT (yellow and red). The NL (green) is docked on the left when the kinesin head has bound ATP. The kinesin head on the right has no bound nucleotide and the NL (red) is disordered. Figure adapted from Corden et al. (2017) and based upon Shang et al. (2014). (**c**) Formation of the Cover-Neck bundle (CNB). Interaction of structure of the two motor heads and the neck coiled-coil. The neck linkers of dimeric kinesin-1 and β0s are shown in red. The ribbon diagram shows the formation of CNB for the plus-end-directed movement (Geng et al. 2014a; Sack et al. 1997)

communicate through the NL, to chemically and physically gate the mechanochemical cycle (Yildiz et al. 2008; Shastry and Hancock 2011). The NL attains more ordered and extended conformation toward the microtubule "plus" end when kinesin binds MTs and ATP and regresses to a more disordered state when γ-phosphate is released after nucleotide hydrolysis (Fig. 15.3b) (Shang et al. 2014).

2-N-Terminal Nonmotor Extension

An earlier study has shown that the presence of N-terminal amino acid residues upstream to the motor domain of kinesin-1 (termed as Cover strand-CS or β0) contributes to the formation of β-sheet upon interaction with the NL region during the plus-end directed motility. This interaction of CS and NL is also called as cover-neck bundle (CNB), essential for the force generation for the plus-end directed molecular motors (Goulet et al. 2012; Geng et al. 2014b, a; Hwang et al. 2008; Khalil et al. 2008). The ATP binding induces the formation of CNB, and further stabilized the motor domain docked conformation for the force generation and plus-end directional motility (discussion in later Sect. 15.2 also). Even though the NL and N-terminal non-motor extension are at opposite ends of the motor core in the primary sequence, they are in close vicinity in the tertiary structure when kinesin moves along MT (Fig. 15.3c). The ribbon diagrams also depict how the NL of the leading head must be unfolded to connect to the trailing head at the beginning of the coiled-coil tail.

15.1.7 Kinesin: Mechanochemistry

Kinesin molecules move on the microtubule surface lattice, interacting with one binding site per tubulin dimer and by doing so, they hydrolyze exactly one adenosine triphosphate (ATP) molecule per 8-nm step (Coy et al. 1994). The transport of cargo by kinesin along the MT has been proposed by two mechanisms, the first one is "hand-over-hand" mechanism, where the kinesin heads step past one another, alternating the lead position (Yildiz et al. 2004), and the second is "inchworm" mechanism, where one kinesin head always leads, moving forward a step before the trailing head catches up (Hua et al. 2002). Although, both the mechanisms are well accepted, the former one is more broadly established (Fig. 15.4c). The unidirectional, i.e., plus end, movement and processivity of kinesin-1 depends on the communication between the two adjoint motors, and it has been proposed that usually these motors walk for hundreds of steps before dissociation.

The nucleotide cycle plays a critical role in the kinesin mechanochemistry and its coupling with the motor domain affects the binding of kinesin with MTs. The binding of ATP to kinesin promotes its high affinity to MT as compared to ADP-kinesin to MTs. The ADP-kinesin binding to MTs augments its dissociation from MTs to complete the cycle (Fig. 15.4a). The ATP binding triggers large conformational changes within the kinesin structural elements leading to the mechanical step and ultimately force generation. During ATP binding, the NL firmly docks toward the catalytic core of the motor domain. After phosphate release, the NL

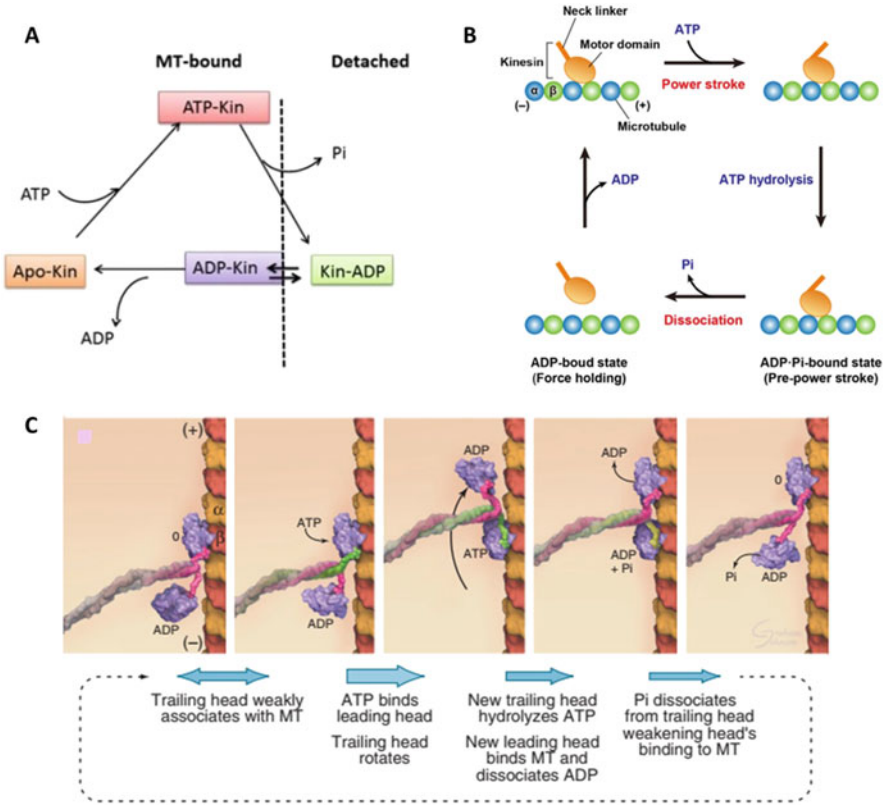


Fig. 15.4 Kinesin-1 ATPase mechanochemistry. (**a**) The schematic diagram of kinesin association with MT coupled with its nucleotide cycle. Figure adapted from Cao (2016). (**b**) The effect of nucleotides and MTs on kinesin-1 mechanochemical cycle. The cartoon diagram depicts the change of different conformations of the NL (orange), as well as binding and dissociation of MTs to kinesin in the presence of ATP and ADP, respectively. The MTs consist of α-tubulin (α) and β-tubulin (β) heterodimer subunits. During power stroke generation, the NL orients from the minus-end direction (left) along an MT to the plus-end direction (right). Figure adapted from Kato et al. (2018). (**c**) The cartoon diagram of "Hand-over-hand" processive stepping mechanism of kinesin dimer along an MT. The docked NL is shown green and undocked NL is pink colored in the cartoon diagram. The figure delineates different steps of the motor movement along an MT tract. Figure adapted from Corden et al. (2017) and based upon Cao et al. (2014)

reverts to its more mobile and disordered conformation that leads to the weaker kinesin-MT binding. Both MT and a γ-phosphate should be present in the active site during the docking of NL and during docking of NL, there is always at least one motor domain attaches tightly to the MT. This condition is observed only with ATP, AMP-PNP (a non-hydrolyzable analog of ATP), and ADP-AlF^{-4}. The NL is detached after the release of ADP. The ATP binding and consequently NL docking lead to the orientation of the trailing motor head toward the plus end of MT that lead to the bias for plus-end-directed motion in the kinesin dimer (Fig. 15.4b, c). The MT

binding not only stimulates kinesin-ADP dissociation but also accelerates ATP hydrolysis. Figure 15.4c depicts step by step the mechanochemical movement of kinesin on MTs in the presence of nucleotides using the "hand-over-hand" mechanism of kinesin-1 stepping.

15.1.8 Mechanism of the Directionality of Kinesin Motors

The N-type kinesin usually moves toward the microtubule plus end, whereas C-type motors move toward the minus end. Kinesin-1, the conventional kinesin is a plus-end-directed motor, whereas *Drosophila* Ncd C-type motor kinesin proteins are minus-end-directed motor (McDonald et al. 1990). Looking at the overall sequence, structures, and even enzymatic activity similarity of these motors, this is quite incredible because they still have opposite directionality behavior (Sablin et al. 1996). Several studies gave insight into the minus-end-directed movement of Ncd. Sablin et al. (1996) demonstrated the difference in the Ncd neck (a coiled-coil) architecture with the equivalent region in the kinesin neck (an interrupted beta-strand), imparts different symmetry to the dimers that are responsible for directional bias on the MT (Sablin et al. 1998). However, Cryo-EM density maps of dimeric motors bound to microtubules delineate a prominent difference between kinesin and Ncd (Hirose et al. 1996; Arnal et al. 1996). The 3D reconstructions of the dimeric motors elucidated that the positioning of the unattached motor toward the MT plus or minus end determines the directionality. The region of the motor responsible for the tilt of the unbound head was thought likely to lie just outside the conserved motor domain. The hypothesis that a region outside the motor domain biases or determines motor directionality was supported by several studies. Constructing chimeric proteins with the Ncd motor domain fused to a kinesin heavy chain (KHC) α-helical coiled-coil stalk and vice versa showed that the stalk–neck region is critical for determining directionality (Henningsen and Schliwa 1997; Case et al. 1997; Endow and Waligora 1998). Alongside, the neck-motor junction, as well as the orientation of neck to stalk region, may also determine directionality for Ncd minus-end directed motors. These studies proposed that the directionality of the kinesin is not solely governed by the motor domain, but the adjacent region near to motor domain may also act as a mechanochemical sensor for directional movement.

15.2 Bidirectional Kinesin-5 Motors

Kinesin-5 are homo-tetrameric bipolar motors and play essential roles in spindle assembly and function by binding to and moving along MTs. They walk along MTs by crosslinking and thereby slide apart the antiparallel spindle MTs, and thus generates the outwardly directed force that separates the mitotic spindle poles during the cell division (Fig. 15.5a, b). Their function well correlates during the mitotic spindle orientation at the midzone where MTs plus-ends are pointing toward the midzone and the bipolar kinesin-5 motors must crosslink and slide antiparallel

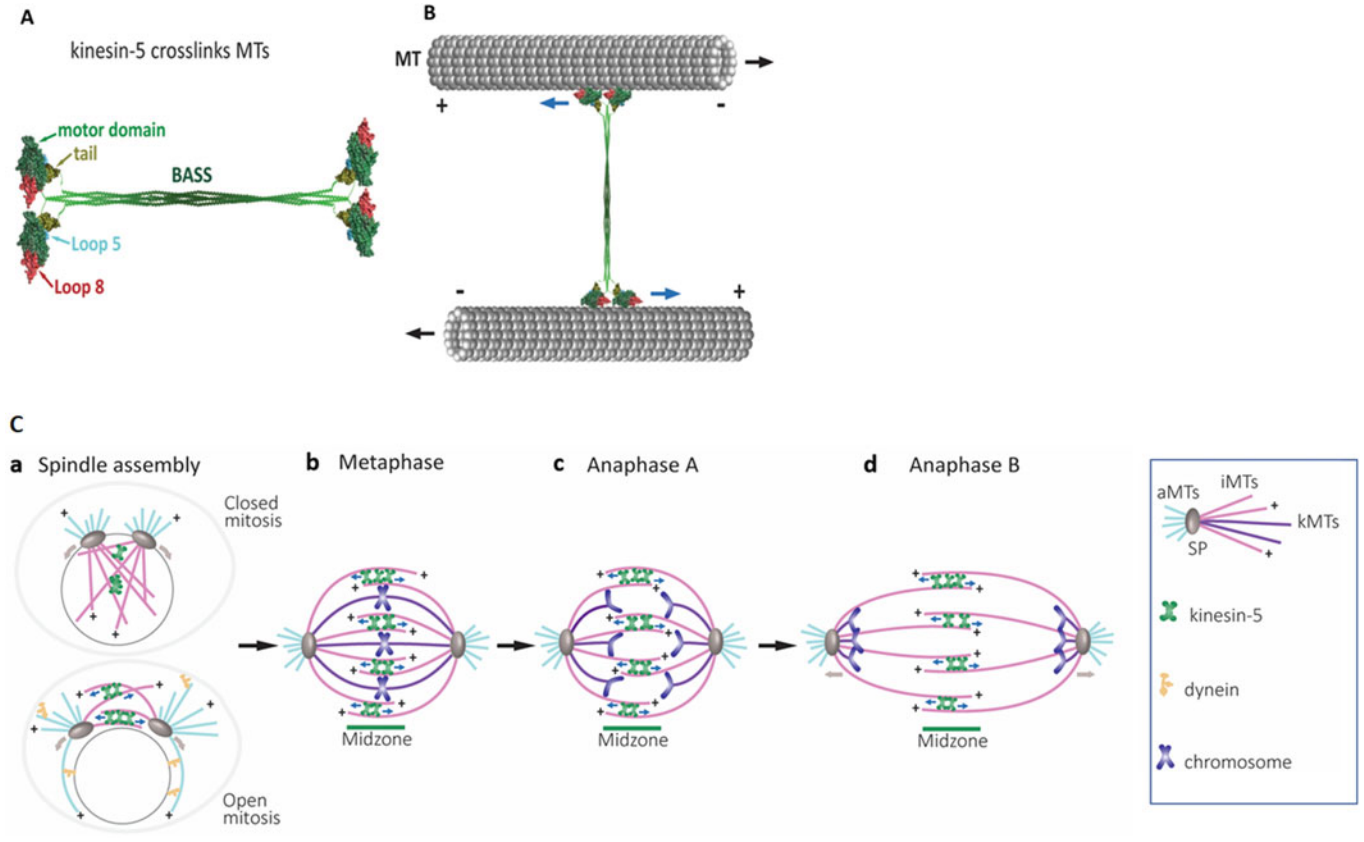


Fig. 15.5 Kinesin-5 plays a major role in mitotic spindle dynamics. (**a**) Schematic representation of a full-length bipolar homo-tetrameric kinesin-5 complex showing the motor domain, coiled-coil BASS domain and tail region (Scholey et al. 2014; Acar et al. 2013). The tetrameric structural homology model of Cin8

was constructed using the Cryo-EM structure of the *S. pombe* Cut7 motor domain (PDB: 5M5I) (Britto et al. 2016). The structural model contains motor domains (green), large Loop 8 (red), Loop 5 in cyan, and the tail domain (olive green). The motor domain and tail domains are connected through the central stalk that includes (BASS) domain necessary for the organization of bipolar homo-tetrameric kinesin-5 complex (Acar et al. 2013). (**b**) The kinesin heads (pair of motor domains) at opposite end interact and crosslink with the two antiparallel MTs and slides them apart. The direction of kinesin-5 head movement on the MTs is shown by blue arrows, and the directions of the MT movement are shown by black arrows. (**c**) The spindle-pole (SP) separation during spindle assembly in closed and open mitosis (Adams and Pringle 1984). The direction of movement of the spindle poles and the kinesin-5 motors are indicated by the brown and blue arrows, respectively. Kinesin-5 motors crosslink antiparallel interpolar MTs (iMTs) at the midzone and stabilize the spindle and further during Anaphase B spindle elongation is marked by the separation of the two opposing spindle poles. Figure adapted from Singh et al. (2018)

interpolar MTs in a plus-end directed manner to separate the spindle poles (Fig. 15.5c). The kinesin-5 motors perform antiparallel sliding only when bound between two antiparallel MTs. They exhibit diffusive bidirectional motility on a single MT (Kapitein et al. 2008).

The homo-tetramer of kinesin-5 motors consists of four identical subunits with two duos of catalytic motor domains on each side of a cylindrical/rod-shaped stalk (Fig. 15.5a) (Cole et al. 1994; Kashina et al. 1996; Gordon and Roof 1999). Some of the important representative members of this family are human Eg5 (HsEg5), *Drosophila melanogaster Klp61F, Xenopus laevis Eg5 (XlEg5) Aspergillus nidulans* BimC, *Schizosaccharomyces pombe* Cut7, and *Saccharomyces cerevisiae* Cin8 and Kip1. The motor domain of kinesin-5 has similarity with other kinesins, but they show very little similarity in the stalk and tail domains. Due to the presence of catalytic ATPase domain at the N-terminus, the kinesin-5 motors were thought to be solely plus-end-directed motors. This was demonstrated in the earlier reports of full-length kinesin-5 proteins from human, *D. melanogaster*, *X. laevis*, and *S. cerevisiae* (Cole et al. 1994; Sawin and Mitchison 1995; Gheber et al. 1999; Kapitein et al. 2005, 2008; Roostalu et al. 2011; Duselder et al. 2012; Fridman et al. 2013). These studies also delineated that the MT-stimulated ATPase rate of Kinesin-5 is slower than that of kinesin-1 indicating differences in the mechanochemical cycle (Rosenfeld et al. 2005; Cochran et al. 2004; Cross 2004). Dimeric human kinesin-5 variant takes approximately eight steps on average before detaching (Valentine et al. 2006), while the dimeric kinesin-1 motors take several hundreds of steps (Vale et al. 1996).

Recently, two independent studies have reported the bidirectional switching of kinesin-5 motors. A study from the Gheber group reported that in in-vitro single-molecule motility assay when moving as a single molecule, *S. cerevisiae* kinesin-5 Cin8 is minus-end-directed under high ionic strength conditions and reverse the directionality in several other experimental conditions (Gerson-Gurwitz et al. 2011). Another group demonstrated that the Cin8 motors switch directionality in response to motor density (Roostalu et al. 2011). Single Cin8 motors preferentially move toward the minus end on the individual MT, whereas the switch to plus-end directed movement when working in a team due to motor–motor coupling while crosslinking the antiparallel MTs. These findings broke a 25-year-old dogma that kinesin motor proteins that carry their catalytic domains at the amino-terminus end move exclusively to the plus-end of the MTs. Furthermore, other kinesin-5 homologs *S. pombe* Cut7 and *S. cerevisiae* Kip1 were also reported to be bidirectional (Fridman et al. 2013; Edamatsu 2014). The steric blockage mechanism has been explicated for Cut7 bidirectional motility as it switches directionality from minus to plus end with crowding by motor and nonmotor proteins (Britto et al. 2016). Thus, three kinesin-5 motors, namely *S. cerevisiae* Cin8 and Kip1 and *S. pombe* Cut7, were shown to be bidirectional in vitro (Gerson-Gurwitz et al. 2011; Roostalu et al. 2011; Fridman et al. 2013; Edamatsu 2014).

15.2.1 Bidirectionality of Kinesin-5: An Insight into Structural Elements

Kinesin-5 motors are homo-tetrameric with pairs of catalytic motor domains located on opposite sides of a 60 nm-long rod-like filament (Kashina et al. 1996; Gordon and Roof 1999; Acar et al. 2013; Scholey et al. 2014) (Fig. 15.5a). This arrangement is structurally opposite to facilitate antiparallel MT sliding to perform their unique mitotic functions. The N-terminal of Kinesin-5 Cin8 contains a nonmotor N-terminal region followed by the catalytic domain, which is followed by a flexible 14–18-amino-acid-long neck linker. The motor domain is further followed by the stalk and tail domain (Hildebrandt et al. 2006). Several studies revealed that the intramolecular domains of kinesin-5 motors such as Loop5 (L5) and Loop8 (L8) region found in the catalytic domain significantly influence the bidirectional behavior of kinesin-5 motors (Behnke-Parks et al. 2011; Bell et al. 2017; Gerson-Gurwitz et al. 2011; Shapira and Gheber 2016). Different structural features influencing the bidirectional motility of kinesin-5 Cin8 have been shown in the structural model as well as in the multiple sequence alignment of various kinesin-5 proteins in comparison with kinesin-1 (Fig. 15.6a, b). The role of several structural domains influencing the directional motility is discussed in the following.

15.2.1.1 N-Terminal Nonmotor Extension and Neck Linker

The NL region is present just after the catalytic domain with 14–18 long amino acid residues. According to the previous structural studies for both kinesin-1 and kinesin-5, this region undergoes conformational alteration upon binding to nucleotides during the motility of kinesins on MTs, i.e., either they dock or undock toward the catalytic motor domain (Shang et al. 2014; Rosenfeld et al. 2001; Turner et al. 2001). Concomitantly, the NL regulates the directional motility and processivity of various kinesin motors (Case et al. 2000; Endow and Higuchi 2000; Endow and Waligora 1998; Rice et al. 1999; Vinogradova et al. 2004; Goulet et al. 2012). The plus-end directional motility is specified by the NL docking of the leading motor domain that positions the trailing motor head forward along the MT track. The kinesin-5 motors may work as an ensemble in comparison to Kinesin-1 due to longer NL residue and hence it was suggested to contribute in relatively reduced processivity than that of kinesin-1 (Duselder et al. 2012; Shastry and Hancock 2011). In kinesin-1, before ATP binding the NL is more flexible or disordered (Rice et al. 1999; Sindelar and Downing 2010; Rosenfeld et al. 2001) and in Kinesin-5, the NL is structured or more ordered and points toward the minus-end of the MTs (Rosenfeld et al. 2005; Turner et al. 2001; Larson et al. 2010; Goulet et al. 2014, 2012). Therefore, the NL conformation and length may substantiate the directional behavior and kinetics of different catalytic stages in kinesin-5 motors in comparison to kinesin-1 motors (Waitzman and Rice 2014; Goulet and Moores 2013; Cochran 2015).

In a plus-end directed motors such as kinesin-1, the docking of NL to the motor domain is stabilized by additional interaction from the N-terminal region termed as CNB that provides the power stroke for force generation by ATP-induced NL docking (discussed in an earlier section) (Hwang et al. 2008; Khalil et al. 2008;

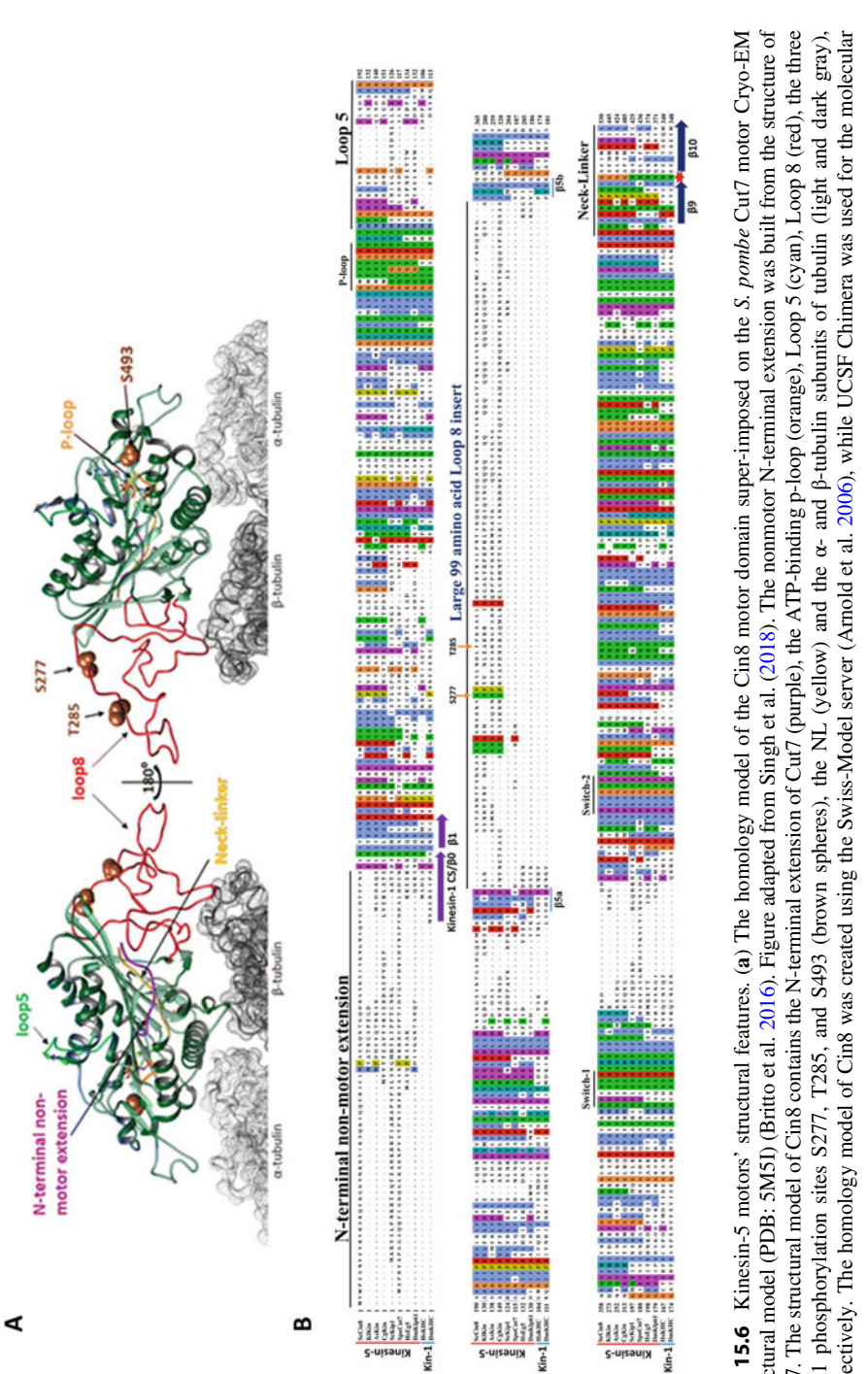


Fig. 15.6 Kinesin-5 motors' structural features. (**a**) The homology model of the Cin8 motor domain super-imposed on the *S. pombe* Cut7 motor Cryo-EM structural model (PDB: 5M5I) (Britto et al. 2016). Figure adapted from Singh et al. (2018). The nonmotor N-terminal extension was built from the structure of Cut7. The structural model of Cin8 contains the N-terminal extension of Cut7 (purple), the ATP-binding p-loop (orange), Loop 5 (cyan), Loop 8 (red), the three Cdk1 phosphorylation sites S277, T285, and S493 (brown spheres), the NL (yellow) and the α- and β-tubulin subunits of tubulin (light and dark gray), respectively. The homology model of Cin8 was created using the Swiss-Model server (Arnold et al. 2006), while UCSF Chimera was used for the molecular

graphics and superimposition (Pettersen et al. 2004). Figure adapted from Singh et al. (2018). (**b**) Multiple sequence alignments of different kinesin homologs, in which top eight sequences are kinesin-5 homologs and bottom two sequences are two kinesin-1 homologs. Organisms with kinesin homolog are indicated on the left: ScCin8: *Saccharomyces cerevisiae* Cin8, KlKin: *Kluyveromyces lactis* kinesin-5, AsKin: *Ashbya gossypii* kinesin-5, *CgKin: Candida Glabrata* kinesin-5, ScKip1: *Saccharomyces cerevisiae* Kip1, SpoCut7: *Schizosaccharomyces pombe* Cut7, HsEg5: *Homo sapiens* Eg5, Dm Klp61F: *Drosophila melanogaster* kinesin-5, HsKHC: *Homo sapiens kinesin*-1, DmKHC *Drosophila melanogaster* kinesin-1. The alignment was done using Unipro UGENE software. The CS/β0-strand (purple) of kinesin-1 N-terminus that can form a CNB while interacting with the docked NL of β9-β10 (blue) is also mentioned along with asparagine latch residue shown with star symbol (red) present between β9 and β10 NL (Budaitis et al. 2019). The unique Cdk1 sites present in ScCin8 are mentioned with residues S277 and T285

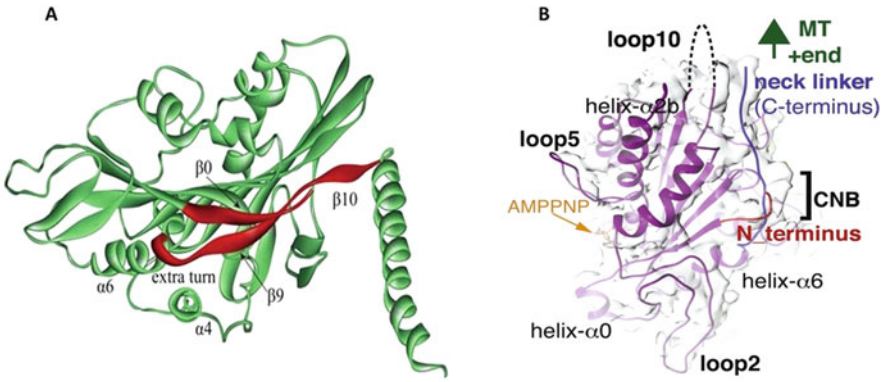


Fig. 15.7 Cover-neck bundle formation (CNB) in kinesin. (**a**) The motor head structure of rat kinesin-1 (PDB ID: 2KIN) in the ATP state conformation showing the β9 and β10 NL and β0 forms a CNB structure (red) (Geng et al. 2014a; Sack et al. 1997). (**b**) CNB formation in kinesin-5 *Schizosaccharomyces pombe* Cut7 (SpoCut7) shows the docked neck linker (blue) interacts with the N-terminus (red) that is directed toward the MT plus end (EMD-3527). Figure adapted from von Loeffelholz et al. (2019)

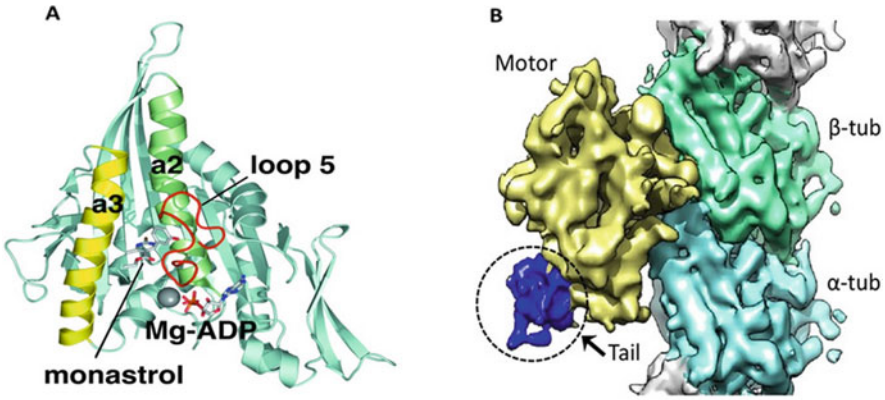
Geng et al. 2014b; Yi-Zhao et al. 2014). The CNB is a 2-stranded β-sheet, formed by the interaction of the β-strand of the NL, β9 with the CS, which is located at the opposite end of the core motor domain (Fig. 15.7a). Afterward, the other C-terminal half of the NL, β10 further docks along with the core motor domain through interactions with α1 and β7. Also, the N-latch residue present between β9 and β10 provides further backbone interaction to complete docking for plus-end directed movement. Any mutation in these residues could weaken CNB formation and kinesin processivity (Budaitis et al. 2019). The kinesin-1 motors are exclusively plus-end directed motors, and they have much shorter nonmotor N-terminal extensions as compared to kinesin-5 (Singh et al. 2018). Recently, the Cryo-EM structure of monomeric bidirectional *S. pombe* kinesin-5 Cut7 shows evidence of CNB formation (Fig. 15.7b) (von Loeffelholz et al. 2019). Therefore, there is a high possibility that in some nucleotide-bound conditions, the kinesin-5 motors may assume different conformations as compared to kinesin-1 that can stabilize the conformation of NL for bidirectional movement. Moreover, the asparagine residue involved in N-latch formation (Budaitis et al. 2019) is highly conserved across processive plus-end directed kinesins except for kinesin-6 family members and some fungal kinesin-5 such as Cin8 where glycine amino acid is present. Therefore, there may be the possibility of different motility properties shown by Cin8 which is due to the difference in residues involved in CNB formation. Further, mutagenesis studies of the N-terminal nonmotor extension, as well as NL residues of bidirectional kinesin-5s, will shed light on the function of this region in the bidirectional motility of kinesin-5 motors.

15.2.1.2 Loop 8

Loop 8 (L8) of kinesin-5 is localized within the β5 strand and separates them as β5a and β5b. Across the kinesin superfamily, L8 shows sequence variation and has been involved in MT binding (Gigant et al. 2013; Kozielski et al. 1997). In kinesin-5 motors, there is high variability in the sequence of L8 from higher eukaryotes to yeast and fungus. Among yeast, *S. cerevisiae* kinesin-5 motor protein Cin8 contains an unusual large 99 amino acid insert in its L8 which is the largest insert among all kinesin motors (Fig. 15.6a, b). The L8 of Cin8 was found to be an important determinant of directionality since the replacement of this non-conserved large insert with the short L8 of the homologous Kip1-induced bias to the minus end directionality of Cin8 in vitro (Gerson-Gurwitz et al. 2011). In a recent study, Bell et al. (2017) demonstrated that L8 is disordered and facilitates the motor domain of Cin8 for noncanonical binding to MTs from usual binding mode indicating the importance of the large L8 of Cin8 in regulating its activity and binding to MTs (Bell et al. 2017). The study delineated that Cin8 L8 may be involved in motor domain oligomerization upon binding the MT lattice and could bias the motor either toward the MT plus-end versus minus-end under different experimental conditions. Therefore, L8 may be involved in the Cin8 cluster formation proposed by Shapira et al. (2017), where they have demonstrated that the kinesin-5 Cin8 switch directionality by clustering and Cin8 multiple tetrameric motors interact through noncovalent interactions (Shapira et al. 2017). The mechanism of cluster formation has been proposed to modulate the motility properties of Cin8 and plays an important role during yeast cell division by the establishment of the bipolar spindle when Cin8 is clustered at the microtubule minus end. Such interactions could be mediated by the large insert in loop 8 of Cin8 (Bell et al. 2017; Gerson-Gurwitz et al. 2011) or by the tail domain (discussion in the next topic) (Duselder et al. 2015). Taking this into account, L8 may act as a molecular switch and regulates the directional behavior of Cin8 (Shapira and Gheber 2016). The presence of three Cdk1 phosphorylation sites in the motor domain of Cin8 (Fig. 15.6a) also regulates the in vivo functions of the protein (Avunie-Masala et al. 2011; Chee and Haase 2010; Goldstein et al. 2017). Among these, two sites (S277 and T285) present in the L8, the S277 site is conserved among fungal kinesin-5 homologs, whereas the T285 site is unique to Cin8 (Fig. 15.6b). It has been demonstrated that the phospho-mimic mutations of Cdk1 sites in L8 affect the binding of Cin8 to MTs sites and promote minus-end directed motility in-vitro (Gerson-Gurwitz et al. 2011; Thiede et al. 2012).

15.2.1.3 Loop 5

Kinesin motors contain a structurally conserved loop near nucleotide-binding pocket (ATP binding site), which is localized between the alpha 2 helices termed as loop 5 (L5). Recent Cryo-EM study showed that L5 acts as a central coordinator of intramolecular rearrangements during the catalytical ATPase cycle in different nucleotide-bound states (Goulet et al. 2012, 2014). It undergoes a change in conformation between "open" and "closed" states and mediates allosteric communication with the nucleotide- and MT-binding sites during the ATPase cycle (Larson et al. 2010; Behnke-Parks et al. 2011; Waitzman et al. 2011; Cochran and Gilbert 2005).



C Table: Steady kinetic parameters for MT-activated ATP hydrolysis

Construct	Source	Ionic Strength	k_{cat} (sec^{-1})	$K_{0.5,MT}$ (nM)
Motor	*Dm KLP61F*	50 mM K Acetate	7.1 ± 0.1	680 ± 48
Motor + Tail	*Dm KLP61F*	50 mM K Acetate	3.5 ± 0.5	757 ± 327
Motor-Tail fusion	*Dm KLP61F*	50 mM K Acetate	3.3 ± 0.1	39 ± 11

Fig. 15.8 (**a**) Eg5 motor domain with monastrol and ADP-magnesium. Ribbon diagram of the Eg5 motor domain (pale blue) with the nucleotide binding containing ADP-magnesium (adapted from PDB: 1Q0B) (Yan et al. 2004). Monastrol bound in a pocket formed by residues in α2 (green, labeled "a2"), α3 (yellow, labeled "a3"), and loop 5 (red). (**b**) Cryo-EM structure of Klp61F motor demonstrates that the motor domain of kinesin-5 interacts with the tail region in the nucleotide-free state. The tail domain is shown in blue, motor domain in yellow, β-tubulin in green, and α-tubulin in cyan. (**c**) Showing the effect of a tail domain on the kinesin-5 motor domain Dm-KLP61F-mediated MT-stimulated ATP hydrolysis. The addition of tail and motor domain in equimolar concentration shows a twofold decrease in k_{cat} (3.5 s^{-1} vs 7.1 s^{-1}) but binding to MT remains almost unaffected K_m (756 nM vs 680 nM) suggesting that MT-stimulated ATPase activity is modulated by tail domain, but motor domain–MT binding remains unperturbed. Figure adapted from Bodrug et al. (2020)

L5 is structurally coupled to both the nucleotide site and the neck-linker element that initiates forward motility in kinesin-family motors possibly using its conformational flexibility to contact the nearby Switch I motif (Behnke-Parks et al. 2011; Larson et al. 2010; Rosenfeld et al. 2005). In addition, deletion or mutating some residues within L5 decreased the MT-stimulated ADP release and nucleotide affinity of kinesin-5 motors that control NL docking as well as directional stepping (Behnke-Parks et al. 2011) (Waitzman et al. 2011; Krzysiak and Gilbert 2006). L5 is unique in kinesin-5 motors as they are a binding site for vertebrate-specific allosteric inhibitors such as monastrol, ispinesib, enastron, and S-trityl-L-cysteine (STLC) that bind to a specific conformation of L5 (Fig. 15.8a) which resulted in inhibition of bipolar spindle formation and mitotic arrest (Kwok et al. 2006; Mayer et al. 1999; Kapoor et al. 2000). Therefore, L5 is considered as a "hot spot" for Eg5 inhibitor binding. These inhibitors bind to a pocket formed by L5, α2 and α3, inhibit ATPase activity by weakening motor–MT interaction and directional kinesin-5 motility (Cochran

et al. 2006; Kwok et al. 2006; Lakamper et al. 2010). The discovery of Eg5 inhibition mediated by L5 prompted the development of new kinesin-5-specific inhibitors for use in anti-cancer treatment (Beer et al. 2008; Tang et al. 2008; Lee et al. 2008; Lad et al. 2008). Intriguingly, L5 *Drosophila melanogaster* kinesin-5 Klp61F and some yeast kinesin-5 are not inhibited by allosteric inhibitors such as monastrol shows the conformation of L5 is rearranged when these kinesin binds to microtubules (Fig. 15.7b) and also the L5 sequence is not well conserved as compared to mammalian kinesin-5 motors (Bodey et al. 2009; Goulet et al. 2012; Maliga and Mitchison 2006; von Loeffelholz et al. 2019). Due to the structure of its L5 insertion, its dynamic structure whose conformation changes through the course of the ATPase cycle (Behnke-Parks et al. 2011) and resistance to mammalian kinesin-5 inhibitors, L5 may be involved in the regulation of the directional motility of yeast kinesin-5 motors.

15.2.1.4 C-Terminal Tail Domain

The directionality preference of *S. cerevisiae* kinesin-5 Cin8 was modulated when the tail domain was truncated (Duselder et al. 2015). Cin8 is minus-end directed in single molecule motility assay under high ionic strength conditions (Gerson-Gurwitz et al. 2011), whereas the tailless construct was slow, bidirectional in motility assay and nonviable in yeast viability assay (Duselder et al. 2015). On the contrary, *S. pombe* Cut7 tail-less construct minus-end directionality was unperturbed (Edamatsu 2014). The importance of tail domain relies on the fact that it was found to crosslink MT in vitro *in S. cerevisiae* Cin8 as well as kinesin-5 motors from *X.laevis* and *D. melanogaster* (Duselder et al. 2015; van den Wildenberg et al. 2008; Weinger et al. 2011). Also, the tail domain of kinesin-1 has been reported to crosslink the two catalytic domains in the active kinesin dimer (Kaan et al. 2011). Very recently, it has been reported that the kinesin-5 tail domain diminishes the MT-stimulated ATP hydrolysis by binding directly to the catalytic motor domain and stabilizing the ADP or nucleotide-free states but does not perturb the MT-motor domain binding (Fig. 15.8) (Bodrug et al. 2020). The above studies show the influence of the tail domain on mammalian kinesin-5 motors. Although there is no report of tail-motor domain interaction in yeast kinesin-5 motors until now, it seems that due to the presence of cdk1 sites for phosphor-regulation and switching of directional motility in the tail-less construct of Cin8, the C-terminal tail may regulate the bidirectional nature of kinesin-5 motors possibly by specific interactions with the catalytic motor domain.

15.3 Conclusion

Kinesins motors are MT-based ATP-powered motors, and perform essential cellular functions. The binding of ATP transcends the small shift in the nucleotide loops to large conformation perturbation that leads to force generation. To date, the mechanism of directionality and different structural elements that control force generation is not fully understood. Different members of kinesin superfamily diversify their functions by rearranging their structural elements to the catalytic motor head domain

in the presence of MTs and energy source ATP. Sequence comparisons of kinesin-5 motors from other kinesin family members have substantiated different structural elements that may be responsible for the bidirectional motility that is unique to yeast kinesin-5 proteins (Singh et al. 2018). The molecular mechanism of force generation, regulation, and the physiological role of the directionality switch of these unique yeast kinesin-5 proteins have not yet been established. In the core of this problem, is the fact that the high-resolution atomic structure of the catalytic dimer has not yet been found that is specifically required for the movement on MTs. However, considering the recent advancement in Cryo-EM and crystallography, the structures of kinesin-microtubule complexes are rapidly improving in resolution, and we can confidently expect the scientific community to elucidate the mechanochemical cycle of kinesin. Concurrently, recent studies with yeast kinesin-5 motors have shown that directional switching is dependent on motor–motor coupling, motor–MT density, ionic strength, and/or parallel versus antiparallel microtubule orientation (Gerson-Gurwitz et al. 2011; Roostalu et al. 2011; Thiede et al. 2012; Fridman et al. 2013). It will be interesting to develop a better understanding of the molecular mechanism that dictates this biophysical conundrum. Aside, the essential role of kinesin-5 motors in mitosis made them an important drug target for antiproliferative agents that bind specifically to mammalian kinesin-5 such as human Eg5. Therefore, there is an urgent need to discover drugs like monastrol, which can selectively inhibit the kinesin-5 motors.

Acknowledgment Dr. Sudhir Kumar Singh would like to thank Ben-Gurion University of the Negev, Israel, for providing Krietman Postdoctoral fellowship to carry out research on the Kinesin-5 motor proteins. We would also like to show our gratitude to Professor Leah Gheber, Department of Chemistry, Ben-Gurion Unversity of the Negev, Israel, for sharing their pearls of wisdom with us during the course of our Postdoctoral research on the Kinesin-5 motor proteins.

References

Acar S, Carlson DB, Budamagunta MS, Yarov-Yarovoy V, Correia JJ, Ninonuevo MR, Jia W, Tao L, Leary JA, Voss JC, Evans JE, Scholey JM (2013) The bipolar assembly domain of the mitotic motor kinesin-5. Nat Commun 4:1343

Adams AE, Pringle JR (1984) Relationship of actin and tubulin distribution to bud growth in wild-type and morphogenetic-mutant Saccharomyces cerevisiae. J Cell Biol 98(3):934–945

Adams RR, Tavares AA, Salzberg A, Bellen HJ, Glover DM (1998) Pavarotti encodes a kinesin-like protein required to organize the central spindle and contractile ring for cytokinesis. Genes Dev 12(10):1483–1494

Allen RD, Metuzals J, Tasaki I, Brady ST, Gilbert SP (1982) Fast axonal transport in squid giant axon. Science 218(4577):1127–1129

Ambrose JC, Li W, Marcus A, Ma H, Cyr R (2005) A minus-end-directed kinesin with plus-end tracking protein activity is involved in spindle morphogenesis. Mol Biol Cell 16(4):1584–1592

Antonio C, Ferby I, Wilhelm H, Jones M, Karsenti E, Nebreda AR, Vernos I (2000) Xkid, a chromokinesin required for chromosome alignment on the metaphase plate. Cell 102 (4):425–435

Arnal I, Metoz F, DeBonis S, Wade RH (1996) Three-dimensional structure of functional motor proteins on microtubules. Curr Biol 6(10):1265–1270

Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22(2):195–201

Avunie-Masala R, Movshovich N, Nissenkorn Y, Gerson-Gurwitz A, Fridman V, Koivomagi M, Loog M, Hoyt MA, Zaritsky A, Gheber L (2011) Phospho-regulation of kinesin-5 during anaphase spindle elongation. J Cell Sci 124(6):873–878

Beer TM, Goldman B, Synold TW, Ryan CW, Vasist LS, Van Veldhuizen PJ Jr, Dakhil SR, Lara PN Jr, Drelichman A, Hussain MH, Crawford ED (2008) Southwest Oncology Group phase II study of ispinesib in androgen-independent prostate cancer previously treated with taxanes. Clin Genitourin Cancer 6(2):103–109

Behnke-Parks WM, Vendome J, Honig B, Maliga Z, Moores C, Rosenfeld SS (2011) Loop L5 acts as a conformational latch in the mitotic kinesin Eg5. J Biol Chem 286(7):5242–5253

Bell KM, Cha HK, Sindelar CV, Cochran JC (2017) The yeast kinesin-5 Cin8 interacts with the microtubule in a noncanonical manner. J Biol Chem 12(797662):797662

Berry RM, Armitage JP (1999) The bacterial flagella motor. Adv Microb Physiol 41:291–337

Bodey AJ, Kikkawa M, Moores CA (2009) 9-Angstrom structure of a microtubule-bound mitotic motor. J Mol Biol 388(2):218–224

Bodrug T, Wilson-Kubalek EM, Nithianantham S, Thompson AF, Alfieri A, Gaska I, Major J, Debs G, Inagaki S, Gutierrez P, Gheber L, McKenney RJ, Sindelar CV, Milligan R, Stumpff J, Rosenfeld SS, Forth ST, Al-Bassam J (2020) The kinesin-5 tail domain directly modulates the mechanochemical cycle of the motor domain for anti-parallel microtubule sliding. Elife 9: e44146

Brady ST, Lasek RJ, Allen RD (1982) Fast axonal transport in extruded axoplasm from squid giant axon. Science 218(4577):1129–1131

Britto M, Goulet A, Rizvi S, von Loeffelholz O, Moores CA, Cross RA (2016) Schizosaccharomyces pombe kinesin-5 switches direction using a steric blocking mechanism. Proc Natl Acad Sci U S A 113(47):E7483–E7489

Budaitis BG, Jariwala S, Reinemann DN, Schimert KI, Scarabelli G, Grant BJ, Sept D, Lang MJ, Verhey KJ (2019) Neck linker docking is critical for Kinesin-1 force generation in cells but at a cost to motor speed and processivity. Elife 8:e44146

Burgess SA, Walker ML, Sakakibara H, Knight PJ, Oiwa K (2003) Dynein structure and power stroke. Nature 421(6924):715–718

Cao L (2016) Structural basis of kinesin motility. Université Paris-Saclay, Saint Aubin

Cao L, Wang W, Jiang Q, Wang C, Knossow M, Gigant B (2014) The structure of apo-kinesin bound to tubulin links the nucleotide cycle to movement. Nat Commun 5:5364

Cao L, Cantos-Fernandes S, Gigant B (2017) The structural switch of nucleotide-free kinesin. Sci Rep 7:42558

Case RB, Pierce DW, Hom-Booher N, Hart CL, Vale RD (1997) The directional preference of kinesin motors is specified by an element outside of the motor catalytic domain. Cell 90 (5):959–966

Case RB, Rice S, Hart CL, Ly B, Vale RD (2000) Role of the kinesin neck linker and catalytic core in microtubule-based motility. Curr Biol 10(3):157–160

Chee MK, Haase SB (2010) B-cyclin/CDKs regulate mitotic spindle assembly by phosphorylating kinesins-5 in budding yeast. PLoS Genet 6(5):e1000935

Cochran JC (2015) Kinesin motor enzymology: chemistry, structure, and physics of nanoscale molecular machines. Biophys Rev 7(3):269–299

Cochran JC, Gilbert SP (2005) ATPase mechanism of Eg5 in the absence of microtubules: insight into microtubule activation and allosteric inhibition by monastrol. Biochemistry 44 (50):16633–16648

Cochran JC, Sontag CA, Maliga Z, Kapoor TM, Correia JJ, Gilbert SP (2004) Mechanistic analysis of the mitotic kinesin Eg5. J Biol Chem 279(37):38861–38870

Cochran JC, Krzysiak TC, Gilbert SP (2006) Pathway of ATP hydrolysis by monomeric kinesin Eg5. Biochemistry 45(40):12334–12344

Cole DG, Saxton WM, Sheehan KB, Scholey JM (1994) A "slow" homotetrameric kinesin-related motor protein purified from Drosophila embryos. J Biol Chem 269(37):22913–22916

Corden JL, Morrison C, Maciver S, Tollervey D (2017) Chapter 36—motor proteins. In: Pollard TD, Earnshaw WC, Lippincott-Schwartz J, Johnson GT (eds) Cell biology, 3rd edn. Elsevier, Amsterdam, pp 623–638

Coy DL, Wagenbach M, Howard J (1994) Kinesin takes one 8-nm step for each ATP that it hydrolyzes. J Biol Chem 274(6):3667–3671

Cross RA (2004) The kinetic mechanism of kinesin. Trends Biochem Sci 29(6):301–309

Demonchy R, Blisnick T, Deprez C, Toutirais G, Loussert C, Marande W, Grellier P, Bastin P, Kohl L (2009) Kinesin 9 family members perform separate functions in the trypanosome flagellum. J Cell Biol 187(5):615–622

Drechsler H, McAinsh AD (2016) Kinesin-12 motors cooperate to suppress microtubule catastrophes and drive the formation of parallel microtubule bundles. Proc Natl Acad Sci U S A 113(12):E1635–E1644

Duselder A, Thiede C, Schmidt CF, Lakamper S (2012) Neck-linker length dependence of processive kinesin-5 motility. J Mol Biol 423(2):159–168

Duselder A, Fridman V, Thiede C, Wiesbaum A, Goldstein A, Klopfenstein DR, Zaitseva O, Janson ME, Gheber L, Schmidt CF (2015) Deletion of the tail domain of the kinesin-5 Cin8 affects its directionality. J Biol Chem 19:620799

Edamatsu M (2014) Bidirectional motility of the fission yeast kinesin-5, cut7. Biochem Biophys Res Commun 446(1):231–234

Endow SA, Higuchi H (2000) A mutant of the motor protein kinesin that moves in both directions on microtubules. Nature 406(6798):913–916

Endow SA, Waligora KW (1998) Determinants of kinesin motor polarity. Science 281 (5380):1200–1202

Endow SA, Kull FJ, Liu H (2010) Kinesins at a glance. J Cell Sci 123(Pt 20):3420–3424

Fridman V, Gerson-Gurwitz A, Shapira O, Movshovich N, Lakamper S, Schmidt CF, Gheber L (2013) Kinesin-5 Kip1 is a bi-directional motor that stabilizes microtubules and tracks their plus-ends in vivo. J Cell Sci 126(18):4147–4159

Geng Y-Z, Ji Q, Liu S-X, Yan S-W (2014a) Initial conformation of kinesin's neck linker. Chin Phys B 23(10):108701

Geng Y-Z, Li T, Ji Q, Yan S (2014b) Simulation study of interactions between kinesin's neck linker and motor domain. Cell Mol Bioeng 7(1):99–105

Gerson-Gurwitz A, Thiede C, Movshovich N, Fridman V, Podolskaya M, Danieli T, Lakamper S, Klopfenstein DR, Schmidt CF, Gheber L (2011) Directionality of individual kinesin-5 Cin8 motors is modulated by loop 8, ionic strength and microtubule geometry. EMBO J 30 (24):4942–4954

Gheber L, Kuo SC, Hoyt MA (1999) Motile properties of the kinesin-related Cin8p spindle motor extracted from Saccharomyces cerevisiae cells. J Biol Chem 274(14):9564–9572

Gigant B, Wang W, Dreier B, Jiang Q, Pecqueur L, Pluckthun A, Wang C, Knossow M (2013) Structure of a kinesin-tubulin complex and implications for kinesin motility. Nat Struct Mol Biol 20(8):1001–1007

Goldstein A, Siegler N, Goldman D, Judah H, Valk E, Koivomagi M, Loog M, Gheber L (2017) Three Cdk1 sites in the kinesin-5 Cin8 catalytic domain coordinate motor localization and activity during anaphase. Cell Mol Life Sci 74(18):3395–3412

Gordon DM, Roof DM (1999) The kinesin-related protein Kip1p of Saccharomyces cerevisiae is bipolar. J Biol Chem 274(40):28779–28786

Goulet A, Moores C (2013) New insights into the mechanism of force generation by kinesin-5 molecular motors. Int Rev Cell Mol Biol 304:419–466

Goulet A, Behnke-Parks WM, Sindelar CV, Major J, Rosenfeld SS, Moores CA (2012) The structural basis of force generation by the mitotic motor kinesin-5. J Biol Chem 287 (53):44654–44666

Goulet A, Major J, Jun Y, Gross SP, Rosenfeld SS, Moores CA (2014) Comprehensive structural model of the mechanochemical cycle of a mitotic motor highlights molecular adaptations in the kinesin family. Proc Natl Acad Sci U S A 111(5):1837–1842

Hardin J (2012) An MBoC favorite: cytokinesis and midzone microtubule organization in Caenorhabditis elegans require the kinesin-like protein ZEN-4. Mol Biol Cell 23(16):3025

Heath CM, Wignall SM (2019) Chromokinesin Kif4 promotes proper anaphase in mouse oocyte meiosis. Mol Biol Cell 30(14):1691–1704

Henningsen U, Schliwa M (1997) Reversal in the direction of movement of a molecular motor. Nature 389(6646):93–96

Hentrich C, Surrey T (2010) Microtubule organization by the antagonistic mitotic motors kinesin-5 and kinesin-14. J Cell Biol 189(3):465–480

Hildebrandt ER, Gheber L, Kingsbury T, Hoyt MA (2006) Homotetrameric form of Cin8p, a Saccharomyces cerevisiae kinesin-5 motor, is essential for its in vivo function. J Biol Chem 281 (36):26004–26013

Hirokawa N (1998) Kinesin and dynein superfamily proteins and the mechanism of organelle transport. Science 279(5350):519–526

Hirokawa N, Noda Y, Tanaka Y, Niwa S (2009) Kinesin superfamily motor proteins and intracellular transport. Nat Rev Mol Cell Biol 10(10):682–696

Hirose K, Lockhart A, Cross RA, Amos LA (1996) Three-dimensional cryoelectron microscopy of dimeric kinesin and ncd motor domains on microtubules. Proc Natl Acad Sci U S A 93 (18):9539–9544

Howard J (2002) Mechanics of motor proteins. In: Flyvbjerg F, Jülicher F, Ormos P, David F (eds) Physics of bio-molecules and cells. Springer, Berlin, pp 69–94

Hu CK, Coughlin M, Field CM, Mitchison TJ (2011) KIF4 regulates midzone length during cytokinesis. Curr Biol 21(10):815–824

Hua W, Chung J, Gelles J (2002) Distinguishing inchworm and hand-over-hand processive kinesin movement by neck rotation measurements. Science 295(5556):844–848

Hwang W, Lang MJ, Karplus M (2008) Force generation in kinesin hinges on cover-neck bundle formation. Structure 16(1):62–71

Janisch KM, McNeely KC, Dardick JM, Lim SH, Dwyer ND (2018) Kinesin-6 KIF20B is required for efficient cytokinetic furrowing and timely abscission in human cells. Mol Biol Cell 29 (2):166–179

Jonckheere AI, Smeitink JA, Rodenburg RJ (2012) Mitochondrial ATP synthase: architecture, function and pathology. J Inherit Metab Dis 35(2):211–225

Kaan HY, Hackney DD, Kozielski F (2011) The structure of the kinesin-1 motor-tail complex reveals the mechanism of autoinhibition. Science 333(6044):883–885

Kapitein LC, Peterman EJ, Kwok BH, Kim JH, Kapoor TM, Schmidt CF (2005) The bipolar mitotic kinesin Eg5 moves on both microtubules that it crosslinks. Nature 435(7038):114–118

Kapitein LC, Kwok BH, Weinger JS, Schmidt CF, Kapoor TM, Peterman EJ (2008) Microtubule cross-linking triggers the directional motility of kinesin-5. J Cell Biol 182(3):421–428

Kapoor TM, Mayer TU, Coughlin ML, Mitchison TJ (2000) Probing spindle assembly mechanisms with monastrol, a small molecule inhibitor of the mitotic kinesin, Eg5. J Cell Biol 150 (5):975–988

Kashina AS, Baskin RJ, Cole DG, Wedaman KP, Saxton WM, Scholey JM (1996) A bipolar kinesin. Nature 379(6562):270–272

Kato Y, Miyakawa T, Tanokura M (2018) Overview of the mechanism of cytoskeletal motors based on structure. Biophys Rev 10(2):571–581

Khalil AS, Appleyard DC, Labno AK, Georges A, Karplus M, Belcher AM, Hwang W, Lang MJ (2008) Kinesin's cover-neck bundle folds forward to generate force. Proc Natl Acad Sci 105 (49):19247–19252

Kim AJ, Endow SA (2000) A kinesin family tree. J Cell Sci 113(Pt 21):3681–3682

Kozielski F, Sack S, Marx A, Thormahlen M, Schonbrunn E, Biou V, Thompson A, Mandelkow EM, Mandelkow E (1997) The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. Cell 91(7):985–994

Kozminski KG, Johnson KA, Forscher P, Rosenbaum JL (1993) A motility in the eukaryotic flagellum unrelated to flagellar beating. Proc Natl Acad Sci U S A 90(12):5519–5523

Kratchmarov R, Kramer T, Greco TM, Taylor MP, Ch'ng TH, Cristea IM, Enquist LW (2013) Glycoproteins gE and gI are required for efficient KIF1A-dependent anterograde axonal transport of alphaherpesvirus particles in neurons. J Virol 87(17):9431–9440

Krzysiak TC, Gilbert SP (2006) Dimeric Eg5 maintains processivity through alternating-site catalysis with rate-limiting ATP hydrolysis. J Biol Chem 281(51):39444–39454

Kull FJ, Endow SA (2013) Force generation by kinesin and myosin cytoskeletal motor proteins. J Cell Sci 126(1):9–19

Kull FJ, Sablin EP, Lau R, Fletterick RJ, Vale RD (1996) Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. Nature 380(6574):550–555

Kwok BH, Kapitein LC, Kim JH, Peterman EJ, Schmidt CF, Kapoor TM (2006) Allosteric inhibition of kinesin-5 modulates its processive directional motility. Nat Chem Biol 2 (9):480–485

Lad L, Luo L, Carson JD, Wood KW, Hartman JJ, Copeland RA, Sakowicz R (2008) Mechanism of inhibition of human KSP by ispinesib. Biochemistry 47(11):3576–3585

Lakamper S, Thiede C, Duselder A, Reiter S, Korneev MJ, Kapitein LC, Peterman EJ, Schmidt CF (2010) The effect of monastrol on the processive motility of a dimeric kinesin-5 head/kinesin-1 stalk chimera. J Mol Biol 399(1):1–8

Larson AG, Naber N, Cooke R, Pate E, Rice SE (2010) The conserved L5 loop establishes the pre-powerstroke conformation of the Kinesin-5 motor, eg5. Biophys J 98(11):2619–2627

Lawrence CJ, Dawe RK, Christie KR, Cleveland DW, Dawson SC, Endow SA, Goldstein LS, Goodson HV, Hirokawa N, Howard J, Malmberg RL, McIntosh JR, Miki H, Mitchison TJ, Okada Y, Reddy AS, Saxton WM, Schliwa M, Scholey JM, Vale RD, Walczak CE, Wordeman L (2004) A standardized kinesin nomenclature. J Cell Biol 167(1):19–22

Lee CW, Belanger K, Rao SC, Petrella TM, Tozer RG, Wood L, Savage KJ, Eisenhauer EA, Synold TW, Wainman N, Seymour L (2008) A phase II study of ispinesib (SB-715992) in patients with metastatic or recurrent malignant melanoma: a National Cancer Institute of Canada Clinical Trials Group trial. Investig New Drugs 26(3):249–255

Levesque AA, Compton DA (2001) The chromokinesin Kid is necessary for chromosome arm orientation and oscillation, but not congression, on mitotic spindles. J Cell Biol 154 (6):1135–1146

Levine C, Hiasa H, Marians KJ (1998) DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities. Biochim Biophys Acta 1400(1):29–43

Liu M, Nadar VC, Kozielski F, Kozlowska M, Yu W, Baas PW (2010) Kinesin-12, a mitotic microtubule-associated motor protein, impacts axonal growth, navigation, and branching. J Neurosci 30(44):14896–14906

Lo KY, Kuzmin A, Unger SM, Petersen JD, Silverman MA (2011) KIF1A is the primary anterograde motor protein required for the axonal transport of dense-core vesicles in cultured hippocampal neurons. Neurosci Lett 491(3):168–173

Lwin KM, Li D, Bretscher A (2016) Kinesin-related Smy1 enhances the Rab-dependent association of myosin-V with secretory cargo. Mol Biol Cell 27(15):2450–2462

Maliga Z, Mitchison TJ (2006) Small-molecule and mutational analysis of allosteric Eg5 inhibition by monastrol. BMC Chem Biol 6:2

Mann BJ, Wadsworth P (2019) Kinesin-5 regulation and function in mitosis. Trends Cell Biol 29 (1):66–79

Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL, Mitchison TJ (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. Science 286(5441):971–974

Mayr MI, Hummer S, Bormann J, Gruner T, Adio S, Woehlke G, Mayer TU (2007) The human kinesin Kif18A is a motile microtubule depolymerase essential for chromosome congression. Curr Biol 17(6):488–498

Mazumdar M, Sung MH, Misteli T (2011) Chromatin maintenance by a molecular motor protein. Nucleus 2(6):591–600

McDonald HB, Stewart RJ, Goldstein LS (1990) The kinesin-like ncd protein of Drosophila is a minus end-directed microtubule motor. Cell 6(6):1159–1165

Miki H, Okada Y, Hirokawa N (2005) Analysis of the kinesin superfamily: insights into structure and function. Trends Cell Biol 15(9):467–476

Minamino T, Imada K, Namba K (2008) Molecular motors of the bacterial flagella. Curr Opin Struct Biol 18(6):693–701

Moores CA, Cooper J, Wagenbach M, Ovechkina Y, Wordeman L, Milligan RA (2006) The role of the kinesin-13 neck in microtubule depolymerization. Cell Cycle 5(16):1812–1815

Muller S, Livanos P (2019) Plant Kinesin-12: Localization heterogeneity and functional implications. Int J Mol Sci 20(17):4213

Nangaku M, Sato-Yoshitake R, Okada Y, Noda Y, Takemura R, Yamazaki H, Hirokawa N (1994) KIF1B, a novel microtubule plus end-directed monomeric motor protein for transport of mitochondria. Cell 79(7):1209–1220

Neupane P, Bhuju S, Thapa N, Bhattarai HK (2019) ATP synthase: structure, function and inhibition. Biomol Concepts 10(1):1–10

Nunes Bastos R, Gandhi SR, Baron RD, Gruneberg U, Nigg EA, Barr FA (2013) Aurora B suppresses microtubule dynamics and limits central spindle size by locally activating KIF4A. J Cell Biol 202(4):605–621

Ogawa T, Nitta R, Okada Y, Hirokawa N (2004) A common mechanism for microtubule destabilizers-M type kinesins stabilize curling of the protofilament using the class-specific neck and loops. Cell 116(4):591–602

Parker J (2000) RNA polymerase. In: Brenner S, Miller JH (eds) Encyclopedia of genetics. Academic, New York, pp 1746–1747

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612

Pollock N, de Hostos EL, Turck CW, Vale RD (1999) Reconstitution of membrane transport powered by a novel dimeric kinesin motor of the Unc104/KIF1A family purified from Dictyostelium. J Cell Biol 147(3):493–506

Popchock AR, Tseng KF, Wang P, Karplus PA, Xiang X, Qiu W (2017) The mitotic kinesin-14 KlpA contains a context-dependent directionality switch. Nat Commun 8:13999

Rayment I, Rypniewski WR, Schmidt-Base K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM (1993) Three-dimensional structure of myosin subfragment-1: a molecular motor. Science 261(5117):50–58

Rice S, Lin AW, Safer D, Hart CL, Naber N, Carragher BO, Cain SM, Pechatnikova E, Wilson-Kubalek EM, Whittaker M, Pate E, Cooke R, Taylor EW, Milligan RA, Vale RD (1999) A structural change in the kinesin motor protein that drives motility. Nature 402(6763):778–784

Roberts AJ, Numata N, Walker ML, Kato YS, Malkova B, Kon T, Ohkura R, Arisaka F, Knight PJ, Sutoh K, Burgess SA (2009) AAA+ Ring and linker swing mechanism in the dynein motor. Cell 136(3):485–495

Roberts AJ, Kon T, Knight PJ, Sutoh K, Burgess SA (2013) Functions and mechanics of dynein motor proteins. Nat Rev Mol Cell Biol 14(11):713–726

Rogers GC, Rogers SL, Schwimmer TA, Ems-McClung SC, Walczak CE, Vale RD, Scholey JM, Sharp DJ (2004) Two mitotic kinesins cooperate to drive sister chromatid separation during anaphase. Nature 427(6972):364–370

Roostalu J, Hentrich C, Bieling P, Telley IA, Schiebel E, Surrey T (2011) Directional switching of the kinesin Cin8 through motor coupling. Science 332(6025):94–99

Rosenfeld SS, Jefferson GM, King PH (2001) ATP reorients the neck linker of kinesin in two sequential steps. J Biol Chem 276(43):40167–40174
Rosenfeld SS, Xing J, Jefferson GM, King PH (2005) Docking and rolling, a model of how the mitotic motor Eg5 works. J Biol Chem 280(42):35684–35695
Sablin EP, Kull FJ, Cooke R, Vale RD, Fletterick RJ (1996) Crystal structure of the motor domain of the kinesin-related motor ncd. Nature 380(6574):555–559
Sablin EP, Case RB, Dai SC, Hart CL, Ruby A, Vale RD, Fletterick RJ (1998) Direction determination in the minus-end-directed kinesin motor ncd. Nature 395(6704):813–816
Sack S, Muller J, Marx A, Thormahlen M, Mandelkow EM, Brady ST, Mandelkow E (1997) X-ray structure of motor and neck domains from rat brain kinesin. Biochemistry 36(51):16155–16165
Sack S, Kull FJ, Mandelkow E (1999) Motor proteins of the kinesin family. Structures, variations, and nucleotide binding sites. Eur J Biochem 262(1):1–11
Sawin KE, Mitchison TJ (1995) Mutations in the kinesin-like protein Eg5 disrupting localization to the mitotic spindle. Proc Natl Acad Sci U S A 92(10):4289–4293
Saxton WM, Hicks J, Goldstein LS, Raff EC (1991) Kinesin heavy chain is essential for viability and neuromuscular functions in Drosophila, but mutants show no defects in mitosis. Cell 64 (6):1093–1102
Scholey JM (2008) Intraflagellar transport motors in cilia: moving along the cell's antenna. J Cell Biol 180(1):23–29
Scholey JM (2013) Kinesin-2: a family of heterotrimeric and homodimeric motors with diverse intracellular transport functions. Annu Rev Cell Dev Biol 29:443–469
Scholey JE, Nithianantham S, Scholey JM, Al-Bassam J (2014) Structural basis for the assembly of the mitotic motor Kinesin-5 into bipolar tetramers. elife 3:e02217
Seidel C, Moreno-Velasquez SD, Riquelme M, Fischer R (2013) Neurospora crassa NKIN2, a kinesin-3 motor, transports early endosomes and is required for polarized growth. Eukaryot Cell 12(7):1020–1032
Shang Z, Zhou K, Xu C, Csencsits R, Cochran JC, Sindelar CV (2014) High-resolution structures of kinesin on microtubules provide a basis for nucleotide-gated force-generation. elife 3:e04686
Shapira O, Gheber L (2016) Motile properties of the bi-directional kinesin-5 Cin8 are affected by phosphorylation in its motor domain. Sci Rep 6:25597
Shapira O, Goldstein A, Al-Bassam J, Gheber L (2017) A potential physiological role for bi-directional motility and motor clustering of mitotic kinesin-5 Cin8 in yeast mitosis. J Cell Sci 130(4):725–734
Shastry S, Hancock WO (2011) Interhead tension determines processivity across diverse N-terminal kinesins. Proc Natl Acad Sci U S A 108(39):16253–16258
Sindelar CV, Downing KH (2010) An atomic-level mechanism for activation of the kinesin molecular motors. Proc Natl Acad Sci U S A 107(9):4111–4116
Singh SK, Pandey H, Al-Bassam J, Gheber L (2018) Bidirectional motility of kinesin-5 motor proteins: structural determinants, cumulative functions and physiological roles. Cell Mol Life Sci 75(10):1757–1771
Snyder JA, McIntosh JR (1976) Biochemistry and physiology of microtubules. Annu Rev Biochem 45:699–720
Sowa Y, Rowe AD, Leake MC, Yakushi T, Homma M, Ishijima A, Berry RM (2005) Direct observation of steps in rotation of the bacterial flagellar motor. Nature 437(7060):916–919
Stumpff J, von Dassow G, Wagenbach M, Asbury C, Wordeman L (2008) The kinesin-8 motor Kif18A suppresses kinetochore movements to control mitotic chromosome alignment. Dev Cell 14(2):252–262
Sutton MD, Walker GC (2001) Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. Proc Natl Acad Sci U S A 98(15):8342–8349
Tang PA, Siu LL, Chen EX, Hotte SJ, Chia S, Schwarz JK, Pond GR, Johnson C, Colevas AD, Synold TW, Vasist LS, Winquist E (2008) Phase II study of ispinesib in recurrent or metastatic squamous cell carcinoma of the head and neck. Investig New Drugs 26(3):257–264

Thiede C, Fridman V, Gerson-Gurwitz A, Gheber L, Schmidt CF (2012) Regulation of bi-directional movement of single kinesin-5 Cin8 molecules. Bioarchitecture 2(2):70–74
Thompson RF, Langford GM (2002) Myosin superfamily evolutionary history. Anat Rec 268 (3):276–289
Tian S, Wu J, Li F, Zou J, Liu Y, Zhou B, Bai Y, Sun MX (2016) NtKRP, a kinesin-12 protein, regulates embryo/seed size and seed germination via involving in cell cycle progression at the G2/M transition. Sci Rep 6:35641
Tokai-Nishizumi N, Ohsugi M, Suzuki E, Yamamoto T (2005) The chromokinesin Kid is required for maintenance of proper metaphase spindle size. Mol Biol Cell 16(11):5455–5463
Tomei EJ, Wolniak SM (2016) Kinesin-2 and kinesin-9 have atypical functions during ciliogenesis in the male gametophyte of Marsilea vestita. BMC Cell Biol 17(1):29
Turner J, Anderson R, Guo J, Beraud C, Fletterick R, Sakowicz R (2001) Crystal structure of the mitotic spindle kinesin Eg5 reveals a novel conformation of the neck-linker. J Biol Chem 276 (27):25496–25502
Vale RD, Reese TS, Sheetz MP (1985) Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. Cell 42(1):39–50
Vale RD, Funatsu T, Pierce DW, Romberg L, Harada Y, Yanagida T (1996) Direct observation of single kinesin molecules moving along microtubules. Nature 380(6573):451–453
Valentine MT, Fordyce PM, Krzysiak TC, Gilbert SP, Block SM (2006) Individual dimers of the mitotic kinesin motor Eg5 step processively and support substantial loads in vitro. Nat Cell Biol 8(5):470–476
van den Wildenberg SM, Tao L, Kapitein LC, Schmidt CF, Scholey JM, Peterman EJ (2008) The homotetrameric kinesin-5 KLP61F preferentially crosslinks microtubules into antiparallel orientations. Curr Biol 18(23):1860–1864
Varga V, Helenius J, Tanaka K, Hyman AA, Tanaka TU, Howard J (2006) Yeast kinesin-8-depolymerizes microtubules in a length-dependent manner. Nat Cell Biol 8(9):957–962
Verhey KJ, Kaul N, Soppina V (2011) Kinesin assembly and movement in cells. Annu Rev Biophys 40:267–288
Vinogradova MV, Reddy VS, Reddy AS, Sablin EP, Fletterick RJ (2004) Crystal structure of kinesin regulated by Ca(2+)-calmodulin. J Biol Chem 279(22):23504–23509
von Loeffelholz O, Pena A, Drummond DR, Cross R, Moores CA (2019) Cryo-EM structure (4.5-A) of yeast kinesin-5-microtubule complex reveals a distinct binding footprint and mechanism of drug resistance. J Mol Biol 431(4):864–872
Waitzman JS, Rice SE (2014) Mechanism and regulation of kinesin-5, an essential motor for the mitotic spindle. Biol Cell 106(1):1–12
Waitzman JS, Larson AG, Cochran JC, Naber N, Cooke R, Jon Kull F, Pate E, Rice SE (2011) The loop 5 element structurally and kinetically coordinates dimers of the human kinesin-5, eg5. Biophys J 101(11):2760–2769
Walczak CE, Verma S, Mitchison TJ (1997) XCTK2: a kinesin-related protein that promotes mitotic spindle assembly in Xenopus laevis egg extracts. J Cell Biol 136(4):859–870
Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1(8):945–951
Wang W, Cantos-Fernandes S, Lv Y, Kuerban H, Ahmad S, Wang C, Gigant B (2017) Insight into microtubule disassembly by kinesin-13s from the structure of Kif2C bound to tubulin. Nat Commun 8(1):70
Weinger JS, Qiu M, Yang G, Kapoor TM (2011) A nonmotor microtubule binding site in kinesin-5 is required for filament crosslinking and sliding. Curr Biol 21(2):154–160
Wolf FW, Hung MS, Wightman B, Way J, Garriga G (1998) vab-8 is a key regulator of posteriorly directed migrations in C. elegans and encodes a novel protein with kinesin motor similarity. Neuron 20(4):655–666

Yan Y, Sardana V, Xu B, Homnick C, Halczenko W, Buser CA, Schaber M, Hartman GD, Huber HE, Kuo LC (2004) Inhibition of a mitotic motor protein: where, how, and conformational consequences. J Mol Biol 335(2):547–554

Yardimci H, van Duffelen M, Mao Y, Rosenfeld SS, Selvin PR (2008) The mitotic kinesin CENP-E is a processive transport motor. Proc Natl Acad Sci U S A 105(16):6016–6021

Yildiz A, Tomishige M, Vale RD, Selvin PR (2004) Kinesin walks hand-over-hand. Science 303 (5658):676–678

Yildiz A, Tomishige M, Gennerich A, Vale RD (2008) Intramolecular strain coordinates kinesin stepping behavior along microtubules. Cell 134(6):1030–1041

Yi-Zhao G, Qing J, Shu-Xia L, Shi-Wei Y (2014) Initial conformation of kinesin's neck linker. Chin Phys B 23(10):108701

Yu KW, Zhong N, Xiao Y, She ZY (2019) Mechanisms of kinesin-7 CENP-E in kinetochore-microtubule capture and chromosome alignment during cell division. Biol Cell 111(6):143–160

Yukawa M, Ikebe C, Toda T (2015) The Msd1-Wdr8-Pkl1 complex anchors microtubule minus ends to fission yeast spindle pole bodies. J Cell Biol 209(4):549–562

Structural and Functional Dynamics of Lysosomal Cysteine Proteases with Particular Reference to Cathepsin B and Cathepsin H

16

Sudhir K. Agarwal, Shalini Singh, and Samir Sharma

Abstract

Cathepsins are a ubiquitously expressed subset of the papain family of 11 distinct cysteine lysosomal proteases, including cathepsins B, C, F, H, K, L, O, S, V, W, and X. They regulate diverse biological processes through protein degradation. Impairment of these processes leads to disease conditions and generates considerable research interest. Cathepsins B and H are fascinating in being functionally different but structurally similar and are involved in neuropeptide and hormone processing. Loss of function has been implicated in various diseases, including cancer. They are also involved in the control of cellular differentiation through regulated proteolysis as well as controlled degradation of the extracellular matrix (ECM). The gain of function in the latter leads to disease states like tumor invasion, metastasis, muscular dystrophy, emphysema, and arthritis. Cathepsin H, uniquely, is an aminopeptidase and endopeptidase, while cathepsin B is an endopeptidase and a carboxypeptidase. This diversity in catalysis and specificity is strongly supported by structure elucidation and folding analysis. This chapter emphasizes the important facts related to cathepsins B and H structure starting from amino acid sequence to gene level and structure–activity relationship of these enzymes. It also attempts to enhance our understanding of cathepsins and presents them as drug targets to control ECM degradation and other components of the disease mechanism.

Keywords

Lysosomal proteases · Cathepsin B · Cathepsin H · Molecular properties · Biosynthesis · Specificity · Structure · Function

S. K. Agarwal (✉) · S. Singh · S. Sharma
Department of Biochemistry, University of Lucknow, Lucknow, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_16

16.1 Introduction

The last few decades are devoted to a group of lysosomal proteases, especially cysteine cathepsins. The reason behind this is the role of thiol cathepsins in various diseases and involvement in intracellular protein degradation. It has been established that these proteases participate in remodeling and degradation of ECM proteins (Buck et al. 1992; Sires et al. 1995; Riese and Chapman 2000; Wolters and Chapman 2000), control of immune response (Riese et al. 1998; Antoniou et al. 2000), tumor metastasis and invasion (Lah and Kos 1998; Coulibaly et al. 1999), and aging alteration (Juhg et al. 1999; Cuervo and dice 2000) in the cell. Intracellular protein degradation occurs in two major sub-cellular systems: lysosomal and non-lysosomal ubiquitin-proteasome system. The lysosomal pathway is the first major site where protein degradation takes place due to action of combined random and limited action of lysosomal proteases containing cysteine cathepsins (Wolters and Chapman 2000). However, non-lysosomal machinery leading the ubiquitin-proteasome system is the other site where most of the endogenous cellular proteins are degraded (Riese and Chapman 2000). Proteolysis is important not only for the renewal of proteins and disposal of defective protein molecules but also for the energetic mobilization of endogenic proteins. Such proteolytic processing can be regulated by the activation of an inactive precursor, accessibility of peptide bond of a substrate, interaction with protease inhibitor, protease specificity, or combination of these factors (Buck et al. 1992). Protein turnover, multiple sclerosis, bacterial and viral diseases, malignancy and muscular dystrophy, among others, are all initiated and/or sustained by well-characterized thiol-dependent cathepsins. These enzymes, particularly cathepsins B and H, thus have no single but multiple functions and have generated massive interest in their properties and structures.

Therefore, the biochemical nature of cysteine cathepsins by which one may be distinguished from the other is briefly focused here. Further, our knowledge of the physiological substrates and inhibitors, structure and mechanism, and function of most of the thiol cathepsins is inadequate as compared to what we know about the regulation, fine structure, and kinetics of other proteases. Structural differences between various cysteine cathepsins result in variations in their substrate specificity and mechanism of inhibition. Although almost all the cysteine cathepsins have been crystallized, information on the amino acid sequences is not available for all and the specificity of these enzymes (Cathepsins B and H) remains controversial. Until now, the role of signal or pro-sequences during the transport/synthesis is not cleared. The present chapter, therefore, emphasizes relatively well-characterized thiol proteases, cathepsins B and H that illustrate the general characteristics in addition to the abovementioned properties and their diversity.

16.2 Lysosomal Cysteine Proteases

The name cathepsin was derived from the Greek word "Kathepsein" meaning to digest. It was first introduced in 1929 by Willstatter and Bamann to describe an acid protease distinct from pepsin. After a decade, Fruton et al. (1941) identified three enzymes in a crude preparation of cathepsin from bovine spleen which were called cathepsins I, II, and III. These enzymes were reclassified in 1952 by Tallan et al., who proposed the name cathepsins A, B, and C that acted on Z-Glu-Tyr, Bz-Arg-NH_2, and Z-Gly-Phe, respectively. Since then, there are about 11 human cysteine proteases, i.e., Cathepsins B, C (J, dipeptidyl peptidase I), F, H(I), K(OC2, O2), L, O, S, V(L2), W(lymphopain), and X(P,Y,Z) (Turk et al. 2001; Rawlings et al. 2010) existing at the sequence level. With the advent of novel concepts and availability of genome sequences (Rossi et al. 2004), this number might probably increase, especially since several new mouse cathepsins without apparent human counterparts have been discovered (Guha and Padh 2008; Vidak et al. 2019). These cathepsins are assigned by simply applying letter designations differing from each other in their distribution, molecular properties, substrate specificity, and sensitivity to inhibitors (Agarwal 1990; Turk et al. 2001). Nomenclature and some molecular properties of lysosomal cysteine cathepsins are summarized in Tables 16.1 and 16.2.

Lysosomal cysteine cathepsins are optimally active at acidic pH values but are unstable at neutral or alkaline pH values. However, cathepsin S is the only exception that retains most of its activity at neutral or slightly alkaline pH (Kirschke et al. 1989). Most of these proteases are glycoprotein in nature, and they are active against large protein substrates and a wide range of small peptides. In 1972, three enzymes isolated from rat liver lysosomes with high proteolytic activity at pH 6–7 were shown to be thiol-dependent; these enzymes are cathepsins B, H, and L (Evered and Whelan 1978). However, two components (B1 and B2) are identified in the same preparation of cathepsin B by Otto (1971), and on the basis of their specificities (McDonald and Ellis 1975) cathepsin B2 is renamed as carboxypeptidase B and cathepsin B1 has been known as cathepsin B. Cathepsin H can readily be distinguished from the other cathepsins by its resistance to high temperatures and by the fact that it possesses both endo- and aminopeptidase activities. Similarly, cathepsin L has been recognized by using Z-Phe-Phe-CHN_2 as a potent inhibitor of the enzyme. Likewise, on the basis of substrate specificity and sensitivity to inhibitors (Tables 16.3 and 16.4), several other thiol cathepsins (Table 16.1) isolated from various sources are found different from each other and cathepsins B, H, and L. However, many of the thiol-dependent cathepsins isolated in relatively small amounts are not yet well characterized. For example, cathepsin K (earlier known as cathepsin N), sometimes called "collagenolytic cathepsin," is analogous to cathepsin L except that it shows slight activity against azocasein (Li et al. 2004). Likewise, beef spleen cathepsin S is similar to cathepsin L, but the two may be differentiated on the pattern of inhibition by Z-Phe-Phe-CHN_2 and capability to hydrolyze synthetic substrates (Barrett and Kirschke 1981). Part of the problem in the study of these enzymes lies in their lack of activity toward commonly used synthetic substrates, which hinders attempts to distinguish one enzyme from the other. Unlike most,

Table 16.1 Nomenclature and tissue expression of human lysosomal cysteine cathepsins

Cathepsin	Synonyms	E.C. number[a]	Tissue expression	Reference
B	Cathepsin B1	3.4.22.1	Widespread	Taniguchi et al. (1985); McDonald and Barrett (1986); Stachowiak et al. (2004)
C	Dipeptidyl peptidase I, cathepsin J	3.4.4.9	Myeloid cells, widespread	BRENDA[b]; McGuire et al. (1997)
F	–	3.4.22.41	Macrophages, widespread	BRENDA[b]; Santamaria et al. (1999)
H	Cathepsin I	3.4.22.16	Widespread	Haraguchi et al. (2003)
K	Cathepsin O, cathepsin OC2,	3.4.22.38	Osteoclasts, bronchial epithelium	Tezuka et al. (1994); Bühling et al. (1999)
L	–	3.4.22.15	Widespread	Barrett and McDonald (1980); Ryvnyak et al. (2004)
O	–	3.4.22.42	Widespread	BRENDA[b]
S	–	3.4.22.27	Antigen-presenting cells	BRENDA[b]
V	Cathepsin L2, cathepsin U	3.4.22.43	Testis, thymic/ coroneal epithelium	Bromme et al. (1999); Tolosa et al. (2003); BRENDA[b]
W	Lymphopain	3.4.22.	CD8+ T cells, NK cells	Wex et al. (1998); Ondr and Pham (2004)
X	Cathepsin Z, cathepsin P, cathepsin Y	3.4.18.1	Widespread	Klemencic et al. (2000)

[a]Nomenclature Committee of the International Union of Biochemistry, in *Enzyme nomenclature,* Academic Press, London (1984)
[b]Taken from database "BRENDA"

cathepsins J (now known as cathepsin C) and K may be differentiated on the basis of their high molecular weights (Liao and Lenney 1984). An unstable protease, now known as cathepsin F, is present in cartilages but resists inhibitors of major classes of proteases like PMSF, IAA, pepstatin, and leupeptin (Barrett and McDonald 1980). However, most of the molecular properties of cathepsins O and W are still unknown.

Thiol reagents include 2-mercaptoethanol, cysteine, cysteamine, dithiothreitol, reduced glutathione, and thioglycerol.

16.3 Biosynthesis and Transport

It is well recognized that lysosomal cysteine cathepsins (B, H, and L) are synthesized on membrane-bound ribosomes, which is supported by the fact that most of these enzymes are glycoproteins, as large precursors are moved cotranslationally into the

Table 16.2 Some molecular properties of lysosomal cysteine cathepsins

Name	Molecular weight (kDa)	Molecular form	pI	Reference
B	25–29	5–7 molecular forms	4.5–5.5	Barrett and Kirschke (1981); Mort and Buttle (1997); Yoshida et al. (2015)
C	230	2 molecular forms	5.8–6.1	Liao and Lenney (1984)
F	34	–	5.2–6.8	Wang et al. (1998)
H	26–28	3 molecular forms	6.0–7.1	Barrett and Kirschke (1981); McDonald and Barrett (1986)
K	650	–	5.3	Liao and Lenney (1984); Drake et al. (1996)
L	23–29	4–6 molecular forms	5.5–6.1	Barrett and Kirschke (1981)
O	–	–	–	Santamaria et al. (1998)
S	14–30	Multiple forms	6.3–7.1	Lautwein et al. (2002)
V	29	–	5.7	Bromme et al. (1999); Yasuda et al. (2004)
W	–	–	–	Wex et al. (1998)
X	32	–	5–5.5	Gunčar et al. (2000)

Golgi apparatus through the lumen of the endoplasmic reticulum (Walter and Blobel 1982; Mainferme et al. 1985; Von Figura and Hasilik 1986; Diment et al. 1988; Smith and Gottesman 1989). During the transfer, it is thought that the precursors are subjected to a number of revisions in their carbohydrate and protein content. The exact site for these modifications is still an enigma. The large precursor has an extra peptide, known as a signal peptide or leader sequence having 15–30 bulky hydrophobic amino acids, is restricted to the NH_2-terminus (Von Figura and Hasilik 1986). At this moment, a complex formed between cytosolic ribonucleoprotein and signal peptide, which is called signal recognition particle (Walter and Blobel 1982), probably regulates the translation of the enzyme. Subsequently, at the surface of the rough endoplasmic reticulum, the complex is attached to a receptor that is known as specific receptor protein (docking protein) (Meyer et al. 1982) and the emerging protein is transported into the lumen of the rough endoplasmic reticulum. The pro-sequences have generally been believed to mediate the localization of the newly synthesized polypeptide chains to their site of action and/or the regulation of their biological activities. Indeed, recovery of enzymatic activity of cathepsin L following the renaturation (Smith and Gottesman 1989) suggests that the propeptide has a crucial role in the folding and/or stability of the enzyme.

Several pieces of evidence suggest that the selective transport of these proteases from the Golgi bodies to the lysosomes is mainly refereed by a receptor situated in the Golgi that identifies mannose-6-phosphate residues/receptors (MPR) which are present in the precursors of lysosomal enzymes (Dingle 1984; Mainferme et al.

Table 16.3 Specific synthetic and protein substrates of lysosomal cysteine cathepsins

Cathepsin	Synthetic substrate	Protein substrate	Reference
B	Z-Arg-Arg-NNap	Insulin, aldolase, collagen and other proteins (pH 3.5–6.0), elastin, kininogen	Brömme et al. (1996); Godat et al. (2004); Yasuda et al. (2004); Lecaille et al. (2007)
C	Z-Phe-Arg-NMec, Gly-Arg-NNap	Gp-96 (Grp 94), calreticulin	McDonald and Barrett (1986); Rawlings and Salvesen (2013)
F	Z-FR-MCA, Z-RR-MCA, Z-Phe-Arg-AMC	Cartilage proteoglycan (pH 4.5)	Barrett and McDonald (1980); Wang et al. (1998); Fonovic et al. (2004)
H	Arg-NNap, Leu-NNap, Bz-Arg-NNap	Proteins (pH 6.0) usually azocasein, kininogen, collagen (pH 6.5)	Brömme et al. (1996); Yasuda et al. (2004); Choe et al. (2006)
K	Z-Arg-Arg-NMec	Collagen, osteonectin, elastin, kininogen	Brömme et al. (1996); Bossard et al. (1996); Godat et al. (2004); Yasuda et al. (2004); Lecaille et al. (2007)
L	Z-Phe-Arg-NMec, Z-Lys-$NPhNO_2$	Proteins (pH 5.0) usually azocasein, elastin, kininogen	Godat et al. (2004); Lecaille et al. (2007)
O	–	–	–
S	GRWPPMGLPWEK -(Dnp)-DArg-NH_2(PMGLP)	Hemoglobin (pH 3.5), elastin, kininogen	Brömme et al. (1996); Godat et al. (2004); Yasuda et al. (2004); Lecaille et al. (2007); Lützner and Kalbacher (2008)
V	Z-Phe-Arg-NHMec, Z-Leu-Arg-NHMec, Z-Val-Arg-NHMec	Serum albumin, collagen, elastin	Bromme et al. (1999); Vidak et al. (2019)
W	–	–	–
X	Abz-FRF($4NO_2$)	Propolypeptides (pH 5.0–6.0), fibronectin, laminin	Docherty et al. (1982); Nägler et al. (1999)

1985; Von Figura and Hasilik 1986). Although considerable work has been done on the transport and processing of cathepsin D (Gieselmann et al. 1983; Matha et al. 2006), from many findings observed by various workers, the role of these receptors is presumed to involve in the formation of a complex with newly synthesized cathepsin in Golgi and deliver the enzyme–receptor complex to an intermediate acidified compartment (endosome). After dissociation of the complex in the endosome, MPR is returned to the Golgi apparatus, while newly synthesized lysosomal cathepsins are transferred to the lysosomes. It should be pointed out that before transferring to lysosomes, a small fraction of these proteases (5–15%) is secreted (Kornfeld 1987). How these secreted enzymes are reached to lysosomes, is not clearly understood, but it seems that they are probably recaptured by receptor-mediated endocytosis and redirect to lysosomes via post-Golgi acidified compartment. It is interesting to note that sorting mechanisms are also possible to exist which

Table 16.4 Inhibitors and activators of lysosomal cysteine cathepsins

Cathepsin	Inhibitor	Activator	Reference
B	PCMB, IAM, NEM, E-64, leupeptin, antipain, chymostatin, CA-074 (analogue of E-64)	Thiol reagents and EDTA	Docherty et al. (1982); Yoshida et al. (2015)
C	PHMB, α_1PI, ATP, IAA, cystatins, Ca^{++}	Thiol reagents and Na^+	McDonald and Barrett (1986); Rawlings and Salvesen (2013)
F	E-64, α_2-macroglobulin, ovalbumin	Pepsin	Wang et al. (1998)
H	PCMB, IAM, NEM, E-64, Leu-CH_2Cl, antipain, cystatins, puromycin, Tos-Lys-CH_2Cl	Thiol reagents and EDTA	Kalnitsky et al. (1983); McDonald and Barrett (1986); Verma et al. (2016)
K	Leupeptin, K^+, ATP, PHMB, E-64, heparan sulfate	Thiol reagents, chondroitin sulfate	Liao and Lenney (1984); Bossard et al. (1996); Turk et al. (2012); Verma et al. (2016)
L	PCMB, IAM, IAA, E-64, Z-Phe-Phe-CHN_2, leupeptin, CLIK-148, CLIK-195	Thiol reagents and EDTA	Barrett and Kirschke (1981); Tsuge et al. (1999); Katunuma et al. (1999); Simmons et al. (2005)
O	–	–	–
S	Z-Phe-Ala-CH_2F, PCMB, Leupeptin, CLIK-060, 2,6-bis-trifluoromethylbenzote	Thiol reagents	Barrett and McDonald (1980); Katunuma et al. (1999); Siklos et al. (2015)
V	Trans-epoxysuccinyl-L-leucylamido-(4-guanidino) butane, cystatin, leupeptin, E-64	Thiol reagents	Santamaria et al. (1998)
W	–	–	–
X	PCMB, antipain, leupeptin	Thiol reagents	Docherty et al. (1992)

do not engage MPR (Dingle 1984; Mainferme et al. 1985). Experimental examinations, however, on the sorting mechanisms for lysosomal cathepsins in the cell are still limited.

Cathepsins B and H, located in the different lysosomes, bind to the cell membrane. However, cathepsin L is situated in the lysosomes which are dispersed diffusely in liver cells. Thus, an important aspect of the functional share of various cathepsins is their different localizations (Ii et al. 1985; Hara et al. 1988; Kominami et al. 1988). Posttranslational processing and maturation of cathepsins B and H given by Katunuma (2010) are summarized in Fig. 16.1. Cathepsin B is translated as 17, 62, and 252 amino acids of prepart, pro-part, and mature part, respectively (Towatari and Katunuma 1978). After removing prepart cotranslationally, the procathepsins are translocated into Golgi apparatus where 38th-Asn in pro-part and 111th-Asn in the mature part are glycosylated by high mannose-type

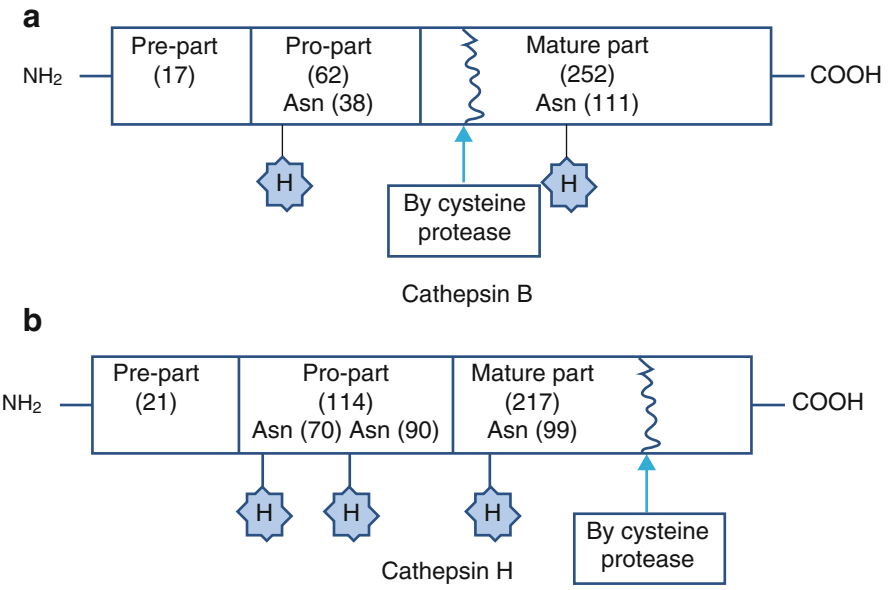


Fig. 16.1 Posttranslational processing and modification of cysteine cathepsins B and H (Kominami et al. 1988)

carbohydrate. Now the mannose-6-phosphate sugars participate as the targeting marker to the lysosomes after cleaving off the pro-part. However, cathepsin H is translated as 21, 114, and 217 amino acids of prepart, pro-part, and mature part, respectively. While the pro-part has two carbohydrate chains at 70th-Asn and 90th-Asn, only one carbohydrate chain is attached at the 99th-Asn in the mature part (Taniguchi et al. 1985; Ishidoh et al. 1987). The commencement of degradation is started from the 47th nicking bond in cathepsin B (Chan et al. 1986) and the 177th nicking bond in cathepsin H (Ishidoh et al. 1987) by some cysteine proteases.

16.4 Properties of Cathepsins B and H

While cathepsin B has been isolated from various sources such as rat liver (Takio et al. 1983), human liver (Barrett and Kirschke 1981; Musil et al. 1991) and placenta (Swanson et al. 1974), bovine spleen (Otto 1971; McDonald and Ellis 1975) and lymph nodes (Zvonar-Popovic et al. 1979), calf brain (Suhar and Marks 1979), rabbit testis (Scott et al. 1987), buffalo liver (Fazili and Qasim 1986; Salahuddin et al. 1996), spleen (Ahmad et al. 1989), kidney (Lamsal et al. 1997) and lung (Agarwal et al. 2016, 2018), porcine spleen (Takahashi et al. 1984a, b), goat spleen (Agarwal and Khan 1987a; Agarwal et al. 1997), and horse muscle (Yoshida et al. 2015), cathepsin H has been purified from human liver (Schwartz and Barrett 1980), kidney (Popovic et al. 1988), brain and meningioma (Chornaya and Lyannaya 2004), rat spleen (Yamamoto et al. 1984) and liver (Kominami et al. 1985), bovine

Table 16.5 Physicochemical properties of buffalo lung cathepsins B and H

Property	Cathepsin B	Cathepsin H
Molecular weight		
SDS-PAGE	23,800	25,400
Gel- filtration	25,400	ND
NH_2-terminal residue	Ala	Lys
COOH-terminal residue	Thr	Val
Isoionic pH	5.12	6.08
Isoelectric pH	4.8–5.2	6.2
Carbohydrate content (%)	4.6	8.7
Tryptophan content [a]	7.8	6.7
SH-group at pH 8.0[a]	0.5	ND
SH-group at pH 8.0 containing urea[a]	1.4	ND
Absorption maxima (nm)	278	280
Emission maxima (nm)	338	342
Specific extinction coefficient ($E^{1\%}_{1\,cm}$)	16.0	18.6

[a]Calculated as moles per mole of the protein
ND Not determined

spleen (Willenbrock and Brocklehurst 1985) and brain (Azaryan and Galoyan 1987), rabbit lung (Singh and Kalnitsky 1978), porcine spleen (Takahashi et al. 1984b), goat liver (Ravish and Raghav 2014), and recently by buffalo lung (Singh et al. 2020). Some of the physicochemical properties of buffalo lung cathepsins B and H studied by us are summarized in Table 16.5. In contrast to cathepsin H, which represents a single-chain enzyme molecule in most of the species, characterization of cathepsin B from porcine spleen (Takahashi et al. 1986a) and goat spleen (Choudhury et al. 1997) shows the presence of two isozymes in these species, suggesting that the cathepsin B isozymes are two separate gene products and/or have a probable tissue/species dependence.

16.4.1 Tissue Distributions

The levels of cathepsins B and H in various rat tissues and peripheral blood cells determined using a sensitive immunoassay (Katunuma and Kominami 1986) are summarized in Table 16.6. Large differences in the concentrations of lysosomal proteases are observed in various tissues. However, it has been verified by immunohistochemical techniques that the activities and concentrations of these proteases are varied within cell type even in one tissue such as brain and liver. The ratio of the levels of cathepsins B and H in tissues also vary: the brain, stomach, esophagus, skeletal muscle, and adrenal gland contain higher levels of cathepsin B, whereas the lung, skin, and liver contain higher levels of cathepsin H (Kominami et al. 1985). Depending on the tissue, several cathepsins represent significant differences in protein expression levels and ratios, implying that each cathepsin(s) may have

Table 16.6 Distribution of cathepsins B and H in various rat tissues and peripheral blood cells[a]

Tissue/cells	Cathepsin B (ng/mg protein)	Cathepsin H (ng/mg protein)
Kidney	1147 ± 240	1429 ± 360
Spleen	542 ± 122	480 ± 195
Liver	316 ± 56	556 ± 201
Vagina	660 ± 94	683 ± 186
Lung	151 ± 26	543 ± 120
Brain	310 ± 86	27 ± 9
Esophagus	208 ± 58	69 ± 21
Stomach	303 ± 62	82 ± 18
Intestine	187 ± 71	116 ± 35
Adrenal gland	437 ± 66	110 ± 26
Bladder	176 ± 32	206 ± 49
Heart	120 ± 36	82 ± 22
Skeletal muscle	92 ± 29	52 ± 16
Testis	62 ± 22	63 ± 24
Skin	252 ± 82	340 ± 78
Lymphocytes	4.1 ± 1.2	16 ± 3.2
Neutrophils	6.2 ± 1.8	16 ± 4.4
Macrophages resident	194 ± 28	158 ± 36
Macrophages elicited	1259 ± 350	298 ± 44
Erythrocytes	>0.5	0.51 ± 0.11

[a]Katunuma and Kominami (1986)

very specific cellular functions (Zavasnik-Bergant and Turk 2007). While cathepsin H is localized in lysosomes of pancreatic islet A-cells, cathepsin B is contained in those of both A- and B-cells (Watanabe et al. 1988; Uchiyama et al. 1994). This heterogeneity among cysteine proteases in lysosomes may reveal the disparity in metabolic substrates between the two cells.

16.4.2 Storage and Assays

Most of the laboratories are facing a problem to retain the activity of either cathepsin B or H during storage and the optimization of their assay systems. Despite several reports on storage of these enzymes where the activity is lost up to 50% or more (Otto 1971; Barrett and Kirschke 1981), cathepsin B can be stored at 0 °C for a long period in buffer concentration more than 0.1 M, pH 5.0, containing 1 mM EDTA (Khan and Ahmad 1987). Similarly, cathepsin H can be preserved for several months at −20 °C in sodium acetate buffer (0.02 M, pH 4.8) having 1 mM EDTA and 0.02% sodium azide at concentrations more than 1 mg/mL (Singh et al. 2020). As much as 90% activity of these enzymes can be recovered by these methods.

Since the activity of an enzyme depends on both the ionic strength of the buffers and the nature of the buffer components, the maximum catheptic activity can be

achieved at low buffer concentrations, preferably in sodium phosphate buffer (0.02 M, pH 6.5), in comparison to other phosphate buffers (Agarwal and Khan 1987b; Singh et al. 2020). These observations also account for the discrepancy in the values of kinetic parameters (K_m, K_{cat}, and V_{max}) of cathepsins B and H reported from different laboratories where the buffers of higher ionic strengths were used (Otto 1971; Barrett and Kirschke 1981; Takahashi et al. 1984a, b; Fazili and Qasim 1986). Although the exact mechanism of buffer ions with the amino acids involved in the active site of the enzyme is not known at present, due care should be taken during the choice of buffers for the assay of lysosomal cysteine proteases.

16.4.3 Enzymes' Nature

Most of the evidence suggests that cathepsins B and H require the integrity of their lone thiol group for the expression of their biological activity. The thiol nature of each enzyme is inferred from its inactivation by stoichiometric amounts of heavy metal ions and thiol-blocking reagents such as IAM, IAA, PHMB, PCMB, and NEM (Otto 1971; Barrett and McDonald 1980; McDonald and Barrett 1986; Agarwal et al. 1997; Singh et al. 2020). Cathepsin B is also inhibited by serum proteins (α_2-macroglobulin, IgG, haptoglobin) (Starkey and Barrett 1973; Barrett and Kirschke 1981), endogenous protease inhibitors (cystatins, stefins, kininogens) (Lenney et al. 1979; Katunuma and Kominami 1986; Turk et al. 2012), low-molecular-weight substances (leupeptin, chymostatin, elastatinal, antipain, E-64) (Takahashi et al. 1984b; Yamamoto et al. 1984; Agarwal et al. 1997; Lamsal et al. 1997), and C-Ha-*ras* gene products (Hiwasa et al. 1987). Likewise, cathepsin H is also inhibited by α_2-macroglobulin (Mason 1989) and by the three groups (stefins, cystatins, and kininogens) of intracellular and extracellular protein inhibitors (Machleidt et al. 1986; Abrahamson et al. 1991; Lenarčič et al. 1996). The enzyme binds with these inhibitors more strongly in comparison to cathepsin B (Guncar et al. 1998). In contrast to cathepsin B and other cysteine cathepsins (L, P, S, and K), cathepsin H is only poorly inhibited by antipain, chymostatin, IAA, mercuric chloride, and by irreversible epoxysuccinyl-based inhibitors derived from E-64 (Barrett et al. 1982; Guncar et al. 1998). Since the selectivity and potency of inhibitors are due to their affinity for the specificity sites of enzyme, the discrepancies in inhibitory effects of these compounds on cathepsins B and H indicate the structural difference of these two enzymes. Further cathepsin H is not or less inhibited by leupeptin (Schwartz and Barrett 1980; Singh et al. 2020), and it is, however, powerfully inhibited by substrate analogues composed of a single amino acid residue bound to a diazomethane or fluoromethane group, which react with the active-site cysteine (Angliker et al. 1989). Although the mechanism of inhibition for a number of thiol protease inhibitors is known, it may, however, be possible that some of these inhibitors (whose mechanism is not known) act by mechanisms analogous to those proposed for the action of serine protease inhibitors.

16.4.4 Specificity

Cysteine cathepsins display broad specificity, splitting their substrates preferentially after basic or hydrophobic amino acid residues. This is factual not only for synthetic substrates but also for protein substrates and steady with their roles in intracellular protein degradation (Turk et al. 2000). These thiol-dependent cathepsins show mainly endopeptidase activity except cathepsins C and X, which are exopeptidases only. While cathepsin C is an aminodipeptidase (Turk et al. 2001), cathepsin X is a carboxymonopeptidase (Gunčar et al. 2000). However, cathepsin B shows an unusual property with its specificity or action on various peptide and protein substrates. The protease is generally assumed to be an endopeptidase because it hydrolyzes amide substrates in which the COOH termini are substituted (Barrett 1977; Bond and Butler 1987). It also possesses the activity of a carboxypeptidase which releases dipeptides sequentially (McDonald and Ellis 1975; Barrett 1977; McKay et al. 1983). However, there is no evidence of specificity for basic amino acid residues (Arg-Arg) like synthetic substrates, if insulin B chain, STI, glucagon, or fructose 1,6-bisphosphate aldolase and many other protein substrates are digested with cathepsin B. Although all the cysteine proteases studied to date exhibit endopeptidase activity toward protein and polypeptide substrates, the display of both endo- and exopeptidase activities, depending on the substrate, by cathepsin B seems to be unusual but not unique; cathepsin H, for example, shows endo- and aminopeptidase activity on polypeptide substrates (Barrett and Kirschke 1981; Koga et al. 1992). The dual activities of the enzyme appear compatible with the well-known specificity of cathepsin H hydrolysis of small synthetic substrates. The enzyme can hydrolyze amide bonds of a substrate with a free α-amino group, for instance Arg-NNap. It can also hydrolyze similar substrates with a blocked α-amino group like Bz-Arg-NNap (Barrett and Kirschke 1981). However, the studies by Takahashi et al. (1988) on porcine spleen cathepsin H show that peptide substrates are cleaved exclusively by aminopeptidase activity; it does not hydrolyze large polypeptides or proteins and thus possesses no detectable endopeptidase activity. Similarly, cathepsin B isolated from either porcine spleen (Takahashi et al. 1986b) or goat spleen (Agarwal 1988) is exclusively a dipeptidyl carboxypeptidase with peptide and protein substrates and has no significant endopeptidase activity.

The assumption that cathepsins B and H are endopeptidases is based on their hydrolyzing activity against synthetic substrates which are hydrolyzed by other endopeptidases such as trypsin and papain. In earlier studies (Barrett 1977; Evered and Whelan 1978; Bond and Butler 1987; Brocklehurst et al. 1987), it was believed that the degradation of proteins by cathepsin B or H came from its endopeptidase activity, though this had never been demonstrated directly. Besides, other thiol cathepsins like L, S, and P in mammalian tissue are similar to cathepsin B or H in many physical properties, and some have overlapping enzymatic activities. Thus, they are not easy to eliminate in a purified cathepsins B and H preparation, and this may account for the apparent endopeptidase activity in some enzyme preparations. Hence, the action of proteases on native protein substrates cannot always be expected on the basis of their action on synthetic substrates. However, the

observation that cathepsins B and H have "no endopeptidase activity" may be a consequence of the substrates tested or species/tissue disparity because the endopeptidase activity of human, bovine, and rat liver cathepsin B is well recognized (Barrett 1977; Bond and Butler 1987; McKay et al. 1983). But the substrate, oxidized insulin B chain, cleaved endoproteolytically by human liver cathepsin B (McKay et al. 1983) cannot be hydrolyzed to any extent by porcine spleen cathepsins B (Takahashi et al. 1986b) or H (Takahashi et al. 1988). Such drastic differences raise the question of whether they are indeed different enzymes. The species dependence of cathepsin B, however, has been confirmed after observing the differences in catalytic and molecular properties of buffalo, bovine, and goat versions of cathepsin B (Agarwal 1991; Lamsal et al. 1997).

Cezari et al. (2002) have inspected the specificity of subsites S_1, S_2, S_1', and S_2' for the carboxydipeptidase activity of cathepsin B with internally quenched fluorescent peptides. Subsite S_1 preferentially accepts basic amino acids for hydrolysis, though substrates with Phe or amino acids bearing an aliphatic side chain at P_1. Despite the existence of Glu^{245} at S_2, the subsite has a clear choice for aromatic amino acid residues, and a substrate with Lys residue at P_2 is hydrolyzed better in comparison to one having an Arg residue. S_1' is a hydrophobic subsite and S_2' exhibits a preference for Phe or Trp residues. In the case of cathepsin H, Takahashi et al. (1988) examined aminopeptidase activity with oligopeptide substrates and suggested that the specificity of the enzyme depends primarily on S_1-side-chain identification. Preferentially released NH_2-terminal residues have large basic (Arg and Lys) or hydrophobic (Phe, Trp, Leu, and Tyr) side chains, whereas the presence of a free α-amino group of substrates is projected to be of secondary significance (Takahashi et al. 1988).

Among the synthetic substrates, the low-molecular-weight peptides containing arginine in the P_1 position (Bz-Arg-NNap, Bz-Arg-$NPhNO_2$) are hydrolyzed by cathepsins B and H very efficiently. McDonald and Ellis (1975), however, have shown that substrates containing paired arginine residues are hydrolyzed most effectively only by cathepsin B. If the N-terminus is unsubstituted, as in Arg-Arg-NNap, then the rate is reduced to about 1% of that with the blocked substrate (Z-Arg-Arg-NNap). The most sensitive substrates discovered for cathepsin H are Arg-NNap and Leu-NNap; Leu-NNap is somewhat less susceptible than Arg-NNap (Barrett and Kirschke 1981). However, a number of other synthetic substrates having different leaving groups such as NNapOMe and NMec have also been accounted for (MacGregor et al. 1979; Barrett and McDonald 1980). The NMec derivatives provide a sensitive fluorimetric assay and are often preferred as they constitute less of a health hazard. The more convenient chromogenic and fluorogenic substrates containing the leaving group 7-amino-4-trifluoromethyl coumarin have been worked out for cathepsins B and H (Tchoupé et al. 1991), but there is still no specific substrate for cathepsin L.

In the absence of the mini-chain, substrate specificity in human cathepsin H shifts from aminopeptidase to endopeptidase (Dodt and Reichwein 2003). This is shown by a genetically engineered mutant of human cathepsin H missing the mini-chain, des[Glu(−18)-Thr(−11)]-cathepsin H, which displays endopeptidase activity

toward the synthetic substrates; it is not cleaved by wild-type recombinant cathepsin H. Nevertheless, the mutant enzyme does not show significant aminopeptidase activity against H-Arg-NH-Mec, a well-known substrate for native human cathepsin H (Dodt and Reichwein 2003). This has been confirmed earlier for cathepsin B in which kinetic data on the interaction of substrates and inhibitors with recombinant variants support the functional activity of the occluding loop of cathepsin B as the main structural part for determining the exopeptidase activity of the protease (Hasnain et al. 1992; Illy et al. 1997; Nagler et al. 1997). Thus, the kinetic studies on substrate hydrolysis and enzyme inhibition reveal the importance of the mini-chain/occluding loop not only as a structural barrier for endopeptidase-like substrate cleavage but also as a structural framework for transition state stabilization of substrates.

16.5 Structure of Cathepsins B and H

From the amino acid sequences of rat liver (Takio et al. 1983), human liver (Ritonja et al. 1985), bovine spleen (Meloun et al. 1988), and porcine spleen (Takahashi et al. 1984b), cathepsin B represents that the enzyme together with cathepsins H, S, K, and L belong to the papain "superfamily." According to the nucleotide sequences, cathepsin B from man (Chan et al. 1986) or rat (Fong et al. 1986) synthesized as a polypeptide chain containing 339 amino acid residues is manufactured to the mature single-chain molecule of 254 amino acids (Nishimura and Kato 1987). Likewise, the amino acid sequences of rat liver (Takio et al. 1983), mouse macrophages (Lafuse et al. 1995), and human kidney (Ritonja et al. 1988) cathepsin H agree with those deduced from the rat (Ishidoh et al. 1989) and human (Fuchs et al. 1988) cDNA sequences. Active human kidney cathepsin H is composed of 230 amino acid residues, 222 of which form a single chain and 8 residue long mini-chain, which is disulfide-linked to the rest of the enzyme (Ritonja et al. 1988). From these sequences, it is evident that the mini-chain brought into existence from cathepsin H propeptide is established between propeptide residues $Thr^{83}P$ and $Glu^{76}P$ (Guncar et al. 1998). However, in some of the mammalian tissues, active cathepsin B or H is present as a two-chain molecule comprising of the light (4–5 kDa) and heavy (20–22 kDa) polypeptide chains cross-linked covalently through a disulfide bridge (Machleidt et al. 1986). This suggests that like other proteolytic enzymes, cathepsins B and H are also synthesized in a precursor form from which the mature enzyme is produced by the removal of the pro-sequence through several proteolytic cleavages.

16.5.1 Sequence Homology

One of the important features of the amino acid sequences of cathepsins B, H, and L, and plant protease papain is the presence of somewhat high level of identity (31–56%) in the amino-(active-site cysteinyl) and (28–44%) in the carboxy-terminal (active-site histidinyl) regions but quite less (13–30%) in the middle region (Takio et al. 1983; Dufour 1988; Ritonja et al. 1988). As can be seen in Table 16.7, about

Table 16.7 Sequence homology in amino acids of cysteine cathepsins[a]

Region of comparison (rat/human)	1–252 (whole protein)	1–77 (active site cysteinyl region)	78–152 (central region)	153–252 (active site histidinyl region)
RCB/RCH	31.5	31.1	13.6	39.8
RCB/papain	30.2	43.7	13.0	28.4
RCH/papain	40.2	47.9	21.4	42.9
HCH/HCB	30.0	31.1	15.5	35.6
HCL/HCB	29.9	34.2	19.5	31.6
HCL/HCH	45.1	56.1	30.4	43.8
Identical residues[b]	13.8	24.3	9.3	11.1

[a]Identity in percent has been calculated from the sequence alignment of rat (Takio et al. 1983) cathepsins B (RCB) and H (RCH), and human (Ritonja et al. 1988) cathepsins B (HCB), H (HCH), and L (HCL)
[b]Identical residues compared in HCH, HCB, HCL, and papain

24% of the residues are identical in the N-terminal region within all four enzymes. Low degree of identity in the central (residues 78–152 in cathepsin B) region, a sector in which a large number of deletions/insertions are found, is perhaps due to a single large insertion of 28–30 residues during the long process of divergent evolution (Dufour 1988). However, the functional importance of this part is not understood, still in plant protease papain. In the C-terminal regions where the active site histidinyl is present, a low level of identity (28–44%) in comparison to the cysteinyl regions (31–56%) proposes that the amino acid sequences in the surrounding area of active site histidinyl residue may reflect the diverse peptidase specificities of cathepsins B and H (Takio et al. 1983). Further, the sequences of cathepsins H and L are very much close to those of the plant enzyme papain than to that of cathepsin B, indicating that cathepsin B must have deviated from the common ancestral gene long before cathepsins H and L.

16.5.2 Active Site

The amino acid residues, Cys^{25} and, as well as Gln^{19}, Asn^{175}, and Trp^{177}, involved in enzyme catalysis are preserved at/around the active site in all cysteine proteinases. Indeed, cathepsins B, H, and L share the amino acid sequence Asn-Ser-Trp (papain residues 175–177) with actinidin and papain. According to X-ray studies on plant thiol proteases, a hydrogen bond formed between Asn^{175} and the active-site His^{159} is shielded by Trp^{177} from the solvent (Dufour 1988). Taking into consideration, a large amount of resemblance in papain and actinidin active-site residues and the sequence homologies among plant and animal cysteine proteinases, all these five enzymes including cathepsins B, H, and L, certainly have same catalytic mechanisms. Any alteration in the active site may happen in the hydrophobic specificity "pocket," the S_2-subsite. For example, the substitution of Ser^{205} in papain by Glu in cathepsin B results in dissimilar surface characteristics of the S_2-subsite

binding site (Baker 1980); the side chain of Glu^{205} may combine with one arginine side chain of the extremely specific substrate Z-Arg-Arg-NMec (Polgar and Csoma 1987).

Whereas six disulfide bridges are restricted in the NH_2-terminal half of the molecule, only two cysteine residues (Cys^{29} and Cys^{240}) of human cathepsin B are unpaired. Topologically Cys^{29} is the same as the reactive (active-site) cysteine (Cys^{25} in papain) of all other thiol proteases; Cys^{240} situated close to the COOH-terminal half of the polypeptide chain surface is unique to cathepsin B (Musil et al. 1991). However, in addition to a disulfide bond produced between Cys^{205} and Cys^{80}P of the mini-chain, porcine or bovine cathepsin H has three disulfide bridges (Cys^{22}–Cys^{63}, Cys^{56}–Cys^{95}, and Cys^{154}–Cys^{200}) which are topologically equivalent to the disulfide bridges in actinidin (Baker 1980). The active-site cleft of porcine (Guncar et al. 1998) or bovine (Baudys et al. 1991) cathepsin H runs the top of the molecule transversely; the wide ends but thin in the middle of active-site cleft contains catalytic (active-site) residues Cys^{25}, His^{159}, and Gln^{19}. In contrast to all other known structures of cysteine proteases, the imidazole ring of His^{159} in cathepsin H does not make a thiolate-imidazolium ion pair with Cys^{25} (Guncar et al. 1998).

16.5.3 Carbohydrate Moieties

While cathepsin B exists in 5–7 molecular forms having pIs in the range of 4.5–5.5 (Table 16.2), cathepsin H presents only in three molecular forms (pIs in the range 6.0–7.1) differing both in carbohydrate contents (Barrett 1977; Barrett and McDonald 1980; McDonald and Barrett 1986). The oligosaccharide structures of lysosomal cathepsins are asparagine-linked and predominantly high mannose type (Kornfeld and Kornfeld 1980). A single *N*-acetylglucosamine and the fucosylated pentasaccharide are present in a molar ratio of 73:27 in porcine spleen cathepsin B (Takahashi et al. 1984a). In contrast, cathepsin H isolated from the same source shows four high mannose-type oligosaccharides having 6–8 mannose residues (Takahashi et al. 1984b). However, a linear tetrasaccharide and a branched pentasaccharide without fucose (absent in porcine spleen cathepsin B) are reported in rat liver cathepsin B (Taniguchi et al. 1985). Similarly, two (high mannose type) oligosaccharides having 9 and 5 mannose residues are found in rat liver cathepsin H beside the three oligosaccharides present in porcine spleen cathepsin H. In both the cases, although the carbohydrates are linked to Asn 111, the structural differences of asparagine-linked sugar chains reflect species and/or organ specificity of glycoproteins among the rat liver and porcine spleen cathepsins.

16.5.4 Structural Transition

An exclusive feature shared by animal thiol proteases, cathepsins B, H, and L, is that they are freely inactivated at neutral pH (Zvonar-Popovic et al. 1980; Ohtani et al. 1982; Khan et al. 1986; Agarwal and Khan 1987a). In contrast to cathepsin L,

cathepsin B is very sensitive toward pH, urea, and guanidine hydrochloride (Agarwal and Khan 1987a; Ahmad et al. 1989; Khan et al. 1992). Buffalo spleen cathepsin B loses its structure as well as the activity irreversibly at alkaline pH; the inactivation of the enzyme is, however, found reversible at acidic pH (Khan et al. 1992). The activity of goat/buffalo enzyme is lost reversibly at denaturant concentrations which does not cause a major change in its secondary structure, and suggest that the inactivation may be ascribed due to slight perturbation in the surroundings of the amino acid residue(s) at and/or around the active site of the enzyme (Agarwal and Khan 1988; Khan et al. 1992). The inactivation process becomes irreversible at high urea/guanidine hydrochloride concentrations leading to the structural changes in the enzyme. Nevertheless, the existence of a multidomain structure in mammalian cathepsin B is first reported after performing a series of denaturation and renaturation experiments by Agarwal and Khan (1988). An important feature of the unfolding–refolding transition of the goat spleen enzyme is that it is not completely reversible and appears to start at extremely low urea concentration. Surprisingly, cathepsin H purified recently from buffalo lung (Singh et al. 2020) unfolds reversibly in two main stages, having a stable intermediate between its native and fully denatured states (unpublished results). The equilibrium and kinetic intermediates have also been confirmed by in vitro studies of cathepsins B, H, and D (Lah et al. 1984; Agarwal and Khan 1988). Although similar data on the precursor "pro-forms" of these enzymes are not available, the non-reversibility of the unfolding transition of the mature enzymes do suggest a role for the pro-sequences in the folding of lysosomal cathepsins.

16.5.5 Secondary Structure

Explorations on the secondary structure of cysteine cathepsins particularly B, H, and L have remained inconclusive so far. According to Garnier et al. (1978), the method predicts the helical structure in cathepsins B, H, and L, respectively, 14, 24, and 16%, whereas the procedure of Chou and Fasman (1974) shows 27, 35, and 30% helical content, respectively. The interpretations of the circular dichroic spectrum (the most commonly used method for determining protein secondary structure) of cathepsin B have varied from one laboratory to the other (Zvonar-Popovic et al. 1980; Bansal et al. 1981; Dufour 1988; Khan et al. 1992). CD spectrum of bovine enzyme conforms to about 12% α-helix and 31% β-sheet (Zvonar-Popovic et al. 1980), whereas buffalo cathepsin B complies with 26% α-helix and 23% β-structure (Khan et al. 1992). This difference in the secondary structure probably seems to be pH-dependent because 34, 65, and 51% α-helix have been reported, respectively, at pHs 5.6, 7.4, and 10.2 in rabbit cathepsin B (Bansal et al. 1981). Moreover, the same pattern is found in cathepsin L where the inactivation at neutral pH is connected to the loss of helical content (40% at pH 5.8 and 17% at pH 7.0) in the enzyme (Dufour et al. 1988). The reason for such an effect may be due to change in the ionization state of histidine side chains which are chiefly situated in the predicted α-helix regions. Further, the ordered structures in cysteine proteases are well preserved in

the NH_2-terminal and COOH-terminal parts (Dufour 1988). Since the maximum insertions/deletions and substitutions in these proteases emerge in the central region, the most important changes in the secondary structures would have occurred in this part. However, the insertion of 28–30 residues in this region of cathepsin B does not change the overall molecular conformation and conformational organization of active site residues in the enzyme (Dufour 1988).

16.5.6 Gene Structure

Whereas the gene structure of cathepsin H has been worked out from rat (Ishidoh et al. 1989) and murine (Buhling et al. 2011), the structure of cathepsin B has been characterized by a genomic DNA segment encoding mouse (Ferrara et al. 1990) and carp (Tan et al. 2006). The isolated clone (λ32) has all the exons matching to the cDNA sequence except for the leader region. The genomic insert spans 15 kbp consisting nine exons encoding 339 amino acids of mouse preprocathepsin B. However, the gene structure of rat cathepsin H comprises at least 12 exons spanning in total more than 21.5 kbp; cathepsin L gene spans 8.5 kbp and comprises eight exons (Ishidoh et al. 1991). A common characteristic for the gene structure of all examined cysteine cathepsins and aleurain (a thiol protease from aleurone cells) is that intron break points are not established at the joints of the pre-peptide, pro-peptide, and mature enzyme regions. Thus, there is no proof that the gene structure of cathepsin B or H communicates to functional parts.

The number and positions of the introns, however, vary between these cathepsins. For instance, the gene encoding rat cathepsins H and L contain 12 and 8 introns, respectively, whereas the gene structure of mouse cathepsin B represents a minimum of nine introns (Ferrara et al. 1990; Ishidoh et al. 1991). Similarly, in cathepsins B and H genes, five introns break off the two active-site cysteine and histidine residues, instead of two in cathepsin L. Like other cysteine protease genes, the region around the active-site (Cys^{29}) residue (the most conserved region) in cathepsin B is cracked by an intron, but on the contrary with cathepsins H and L the intron break point is positioned immediately following the active site. The differences in both the number and position of introns between thiol protease genes suggest that the relation between the genes is not direct. Since cathepsin H gene is formed of four rather than two ancestral gene parts found in aleurain, and the GC content of dissimilar exons are more uniform for cathepsin B gene than for cathepsins H and L (Ferrara et al. 1990; Ishidoh et al. 1991), the earlier notion that four enzymes (cathepsins B, H, L, and papain) derived from a common ancestral gene seems not to be true. However, the preserved sequence around the cysteinyl active site which has probably evolved in numerous ways in these enzymes proves an important function of this region for hydrolyzing activity of thiol cathepsins.

16.5.7 Crystal Structure

Among the investigated 11 cysteine human cathepsins, cathepsin B is the first one whose crystal structure is determined in the 1990s (Musil et al. 1991) which is followed by cathepsins K (McGrath et al. 1997), L (Fujishima et al. 1997), H (Guncar et al. 1998), V (Somoza et al. 2000), X (Gunčar et al. 2000), C (Turk et al. 2001), F (Somoza et al. 2002), and S (Turkenburg et al. 2002). Although three-dimensional (3D) structure of two more human thiol cathepsins, O and W, is not yet known, a 3D-based sequence arrangement of the mature structure of the nine cysteine cathepsins with identified 3D structure reveals conservation of the active-site residues (Cys^{25} and His^{163}, cathepsin L numbering), the N-terminus Pro^{2}, the residues interact with the main chain of the bound substrate (Gln^{19}, Gly^{68}, and Trp^{183}) and certain cysteine residues (Turk et al. 2012). The plant thiol proteases for which 3D structures deduced earlier by X-ray diffraction data are papain (Drenth et al. 1968; Kamphuis et al. 1984) and actinidin (Baker and Dodson 1980). These enzymes exhibit the same conservation pattern as present in lysosomal cysteine cathepsins.

The crystal structures of porcine cathepsin H (Guncar et al. 1998) and human cathepsin B (Musil et al. 1991) have been deduced by X-ray crystallography at the resolution of 2.1 Å and 2.15 Å, respectively. In each case, the enzyme consists of a single polypeptide chain folded to form two domains (left-hand "L" and right-hand "R") with a deep cleft between them. The L-domain is mainly α-helical having the longest central helix and the R-domain is based on a type of β-barrel facing strand (s) which form a coiled structure; the barrel is encircled by α-helix at the bottom. The two domains interrelate via an extended amphipathic interface stabilized by several hydrogen bonds plus hydrophobic interactions. The interface unlocks at the top into a V-shaped active-site cleft where two catalytic residues, cysteine and histidine, are situated. The reactive site cysteine is located at the N-terminus of the central helix of the L-domain, whereas the histidine is situated within the β-barrel residues of the R-domain. A thiolate-imidazolium ion pair formed between two catalytic residues is essential for the proteolytic activity of cathepsin B and other cysteine cathepsins except for cathepsin H (Guncar et al. 1998; Turk et al. 2012).

Human cathepsin B is roughly disk-shaped having a thickness of 30 Å and a diameter of 50 Å (Musil et al. 1991). Out of 248 distinct cathepsin B amino acid residues, 166 α-carbon atoms are topologically equivalent with α-carbon atoms of papain. But several big insertion loops which modify its properties are present on the molecular face. The occluding loop containing 108–119 amino acid residues of the enzyme takes up the back of the active-site cleft resulting in cathepsin B, a carboxy-peptidase. Seven disulfide connectivities are in full agreement with those determined for bovine cathepsin B by chemical methods (Baudyš et al. 1990). However, porcine cathepsin H is an ellipsoidal molecule having dimensions $32 \times 26.5 \times 24$ Å (Guncar et al. 1998). Superposition of the α-carbon atoms of cathepsin H on the α-carbon atoms of actinidin, papain, and cathepsin B exhibits that 180 amino acids of cathepsin H and actinidin (Baker 1980) are topologically equal, 173 with papain (Kamphuis et al. 1984) and 156 with human cathepsin B (Musil et al. 1991).

However, the mini-chain (an octapeptide linked to the R-domain) attached through a disulfide bridge to Cys^{205} of the body of cathepsin H is implicated in the steric regulation of the accessibility of the active-site cleft (Baudys et al. 1991) specifying aminopeptidase activity of the enzyme (Guncar et al. 1998). Moreover, the structure strengthens the outline of disulfide bridges in bovine cathepsin H (Baudys et al. 1991); the three disulfide bonds in the enzyme are topologically equivalent to the disulfide bridges in actinidin (Baker 1980). The stability of the mini-chain in cathepsin H and procathepsin H is studied recently by Hao et al. (2018), and the results indicate that the mini-chain is indeed more dynamic in procathepsin H, whereas it reorients to the more stable conformation in cathepsin H during the process of activation.

16.6 Overall Structure–Activity Relationship

One of the significant aspects, i.e., specificity diversity, has now been cleared from the crystal structures of cathepsins B and H. The 3D structures displayed that exopeptidase exhibits extra structural features that alter the active-site cleft (Musil et al. 1991; Guncar et al. 1998). While the active-site cleft expands along the whole length of the two-domain interface in endopeptidases, additional features reduce the number of binding sites in exopeptidases (Turk et al. 2003).

In cathepsin B, substrate binding is governed by a novel insertion loop that packs the active-site cleft on the primed subsites and seems to favor binding of peptide substrates with two residues carboxy-terminal to the scissile peptide bond; the occluding loop of the enzyme uses two histidine residues (His^{110} and His^{111}) to port the C-terminal carboxylic group of the peptidyl substrate, suggesting an explanation for the well-known dipeptidyl carboxypeptidase activity of the enzyme (Musil et al. 1991). The other subsites neighboring to the reactive site Cys^{29} are quite similar to papain; Glu^{245} in the S_2 subsite supports the basic P_2 side chain. Besides the histidine residues, the hidden Glu^{171} might stand for a group with pK_a of ~5.5 close to the active site, which controls exo- and endopeptidase activity of the enzyme. Since the exact role of these residues remains speculative, it may be further clarified by recombinant methods.

However, cathepsin H utilizes a region from its propeptide part to represent features that fill up the active-site cleft on the non-primed subsites S_2 and S_3; the enzyme uses a carboxylic group of the main and/or side-chain residues to port the positively charged NH_2-terminus of the peptidyl substrate. This carboxylic group is situated at the COOH-terminus of an octapeptide element (originate from the propeptide), termed the mini-chain which remains linked to the active-site cleft through the side chains of $Gln^{78}P$, $Cys^{80}P$, and $Thr^{83}P$ of cathepsin H after its activation, providing an essential responsibility of mini-chain in the aminopeptidase activity of the enzyme (Guncar et al. 1998). This has, however, been confirmed after production and characterization of cathepsin H lacking mini-chain, resulting in a switch of its substrate specificity to endopeptidase (Dodt and Reichwein 2003; Vasiljeva et al. 2003). Fascinatingly, in aminopeptidases, glycosylation not only

plays a crucial role in stabilizing the structure of the added characteristics but also fills up the active-site cleft tighten the substrate-binding (Turk et al. 2012). Unlike other cysteine proteases, cathepsin H is not inhibited by its own free propeptides (Horn et al. 2005).

16.7 Physiological and Pathological Implications

Lysosomes are the only cellular compartment having a series of hydrolases including cathepsins for the complete degradation of all classes of macromolecules by different modes (Wolters and Chapman 2000; Reiser et al. 2010; Vidak et al. 2019). Although it has not yet been possible to assign a precise function to a particular enzyme, various roles have been proposed for lysosomal thiol-dependent cathepsins. However, pieces of evidence obtained using thiol-specific inhibitors indicate that cysteine cathepsins particularly B, H, and L contribute a significant role in protein turnover (Bohley et al. 1974; Barrett 1977; Evered and Whelan 1978; Brocklehurst et al. 1987; Bromme and Wilson 2011; Verma et al. 2016). For example, the digestion of liver cytosolic proteins by rat liver lysosomal enzymes is entirely due to cathepsin B (Dean 1976) and at pH 6.0 short-lived cytosolic proteins are hydrolyzed in preference to long-lived proteins by cathepsin L (Bohley et al. 1974). In pulmonary emphysema, cathepsin B not only digests lung structural proteins but also inactivates enzymatically α_1-proteinase inhibitor and reduces its protective concentration in and around lung tissues (Gairola et al. 1989). Moreover, cysteine cathepsins like H and K found in the lung are associated with inflammatory lung diseases (Chilosi et al. 2009; Faiz et al. 2013).

Apart from the fact in general protein turnover, the cysteine cathepsins could also have a role in the specific processing of proteins and thus in the regulation of enzymatic activity. Cathepsins B and L inactivate aldolase when tested with fructose-1,6-bisphosphate as a substrate (Bond and Butler 1987). The enzyme-treated aldolase showed no detectable change in molecular weight, suggesting that the modification may be significant for the regulation of aldolase activity. Certain other enzymes like glucokinase, pyruvate kinase, tyrosine- and alanine aminotransferases, asparaginase, and glyceraldehyde 3-phosphate dehydrogenase are also inactivated by cathepsin B (Evered and Whelan 1978; Brocklehurst et al. 1987; Barrett et al. 1998). Likewise, cathepsins B and H can also activate peptide hormones and various proteins by cleavage of their precursor forms, e.g., the conversion of proinsulin to insulin (Docherty et al. 1982), proalbumin to albumin (Quinn and Judah 1978), and trypsinogen to trypsin (Otto and Reisenkonig 1975). Furthermore, these enzymes participate in various other physiological processes such as protein synthesis, growth and aging, fertilization, memory, tissue resorption, and modeling (Bond and Butler 1987; Barrett et al. 1998). The exact mechanism of various inactivations or conversions of different enzymes/proteins by cathepsins B and H or other cysteine cathepsins is, however, still speculative.

It has been widely accepted that the extracellular matrix (ECM) is a reservoir for endogenous growth factors. In the body, endogenous proteases such as matrix

metalloproteinases and their inhibitors are involved in routine ECM turnover for the maintenance of healthy tissue (Docherty et al. 1992; Burgess et al. 2009; Faiz et al. 2013). Among the papain family of proteases, cathepsins are capable of degrading ECM components with unique collagenolytic activity; cysteine cathepsins B, H, F, K, L, and S have the potential to participate in wound healing (Wolters and Chapman 2000) and ECM remodeling (Lutgens et al. 2007). Besides their role in ECM degradation, they are implicated in major histocompatibility complex (MHC) class II molecules which are expressed on the surface of antigen-presenting cells where after binding with exogenous proteins, MHCs present them to CD4+ T cells. Further, these cathepsins are also involved in the development and progression of cardiovascular diseases such as atherosclerosis, aneurysm, cardiac repair, and cardiomyopathy (Cheng et al. 2012). Nonetheless, imbalance in expression between cysteine cathepsins (S, K, L, and B) and their endogenous inhibitors (cystatin C) may favor proteolysis of ECM in the pathogenesis of such cardiovascular diseases (Wu et al. 2018). There is evidence for cathepsin B as a vital drug target for traumatic brain injury in which the enzyme gets away from its usual subcellular location (lysosome) to ECM (cytoplasm) where the unleashed proteolytic control causes devastation via autophagic, necrotic, apoptotic, and activated glia-induced cell death simultaneously with inflammation and ECM breakdown (Hook et al. 2015).

The activity of cathepsins (B, H, and L) is altered in several disease states such as muscular dystrophy, malignancy, ischemia, hypervitaminosis, multiple sclerosis, diabetes, arthritis, and various forms of cancer cells including the invasion of host tissue and metastasis (Evered and Whelan 1978; Brocklehurst et al. 1987; Turk et al. 2001; Vasiljeva et al. 2006; Victor et al. 2011; Verma et al. 2016). The mechanism for the precise regulation of these proteolytic enzymes however remains to be established. A report published in 1981 that cathepsin B activity is significantly elevated in a variant of B_{16} melanoma with high metastatic potential (Sloane et al. 1981), and it seems probable that the enzyme release from tumor cells may facilitate invasion and extravasation of tumors. This has been further confirmed by Weiss et al. (1990), who observed that invasive tumor cells enhance the level of cathepsin B in their plasma membranes, which may be used to degrade basement membrane components such as laminin and thereby facilitate tumor invasion. In human melanoma (primary and metastatic) cell lines, cathepsin B is highly expressed at the surface of metastatic but not of primary melanoma cell lines; chemical (CA-074 and CA-074Me) and biological (specific antibodies) inhibitors exert a powerful anti-invasive activity by a mechanism that brings into play the impairment of metastatic cell dissemination. However, in vivo studies (in murine xenografts), human melanoma growth, and artificial lung metastases are significantly reduced by CA-074, suggesting a role for cysteine protease in tumor growth and metastatic potential of human melanoma (Matarrese et al. 2010). Literally, what is the role of protease inhibitors (endogenous/chemical/biological) in tumor invasion and how they regulate the invasive potential of tumor cells is still unclear. Doxorubicin is an effective cytotoxic anticancer drug used for the treatment of malignancies and a broad range of solid tumors, and it shows severe dose-dependent toxicities (Gianni et al. 2003). However, a number of studies on cancer cells in vitro and tumor xenograft in vivo

have revealed that cathepsin B-cleavable doxorubicin prodrugs are less toxic in vitro and more effective in vivo, suggesting the role of this enzyme-cleavable prodrugs in cancer therapy (Zhong et al. 2013).

Although fewer studies have been done on cathepsin H in cancer, the activity of the enzyme is elevated in breast cancer, melanoma, tumor invasion, colorectal and prostate carcinomas, and tumor vasculature (Gabrijelcic et al. 1992; Kos et al. 1997; del Re et al. 2000; Waghray et al. 2002; Gocheva et al. 2010). However, reduced cathepsin H expression has also been reported in squamous cell carcinomas of the head and neck and mixed expression patterns in pancreatic cancer cells (Kos et al. 1995; Paciucci et al. 1996). The differences in the expression pattern of cathepsin H in various cancers may indicate highly specific functions for the enzyme in different tissues at various stages of cancer. Moreover, the probable role of cathepsin H in tumor progression is its capability to degrade fibrinogen and fibronectin, suggesting that the enzyme may be occupied in the destruction of ECM components leading to cancer proliferation, migration, and metastasis (Tsushima et al. 1991; Turk et al. 2012). Nevertheless, a study on T_3-mediated upregulation of cathepsin H involved not only in extracellular signal-regulated kinase activation but also in increased cell migration reveals that overexpression of cathepsin H in a subset hepatoma is thyroid-hormone-receptor-dependent having a significant role in hepatoma progression (Wu et al. 2011). Apart from cystatins (stefins A, B, and cystatin C), cathepsins B and H have been reported as significant prognostic markers in sera of patients with melanoma and colorectal cancer (Kos and Schweiger 2002).

16.8 Conclusions and Future Perspectives

While cysteine cathepsins like B, H, L, and S show the similarity in terms of physical properties, enzymatic activities, and homology with each other in amino acid sequences including the essential catalytic site region, it is not easy to distinguish these enzymes with respect to their biological functions, and it remains difficult to establish what role they take part in pathophysiological protein degradation. Similarly, a large number of different proteins can act as substrates for cysteine cathepsins in vitro studies, but there is little evidence to confirm that such reactions occur in vivo. Hence the substrate specificity and specific cleavage sites of these cathepsins are smart areas of study for accepting their role in life events and for designing drugs against these enzymes. However, the wide variations in tissue levels of cathepsins B and H are compatible with specific functions of these proteases in distinct tissues. It is, therefore, generally accepted that these enzymes participate in the breakdown of both intra- and extracellular proteins. Despite their several roles in protein catabolism, the exact mechanism of action of each cathepsin is still unknown. How pro-sequences help in the correct folding of a cathepsin molecule is yet to be explained. Moreover, the involvement of cathepsins B and H in inflammatory reactions has been proposed on evidence from inhibition studies and the detection of significant catheptic activities at inflammation sites. However, it is not yet certain that which cysteine cathepsin is directly responsible for the tissue

breakdown at inflammatory sites. Gene structure and/or antibodies may be useful to provide a clue for understanding the molecular evolution and functional diversity among cysteine cathepsins.

Cysteine cathepsins are also emerging as major players in tumor progression, making them potential drug targets for a wide range of human cancers. Although cathepsins B and H have been used as drug targets to control ECM degradation and various metabolic activities involved during disease progression, the use of cysteine protease inhibitors (either endogenous, chemical, or biological) may be taken as a pioneering approach in the management of metastatic melanoma as well as on other carcinomas before making therapeutic strategies. Endogenous intracellular inhibitors are likely important in the control of these proteases, and the fluctuations in enzyme activities in cells are due to changes in inhibitor, rather than protease concentrations. Future studies in both clinical samples and preclinical models should now allow us to find out whether these cathepsins have similar or unique roles in different tumor microenvironments. Furthermore, these enzymes have been targeted by pharmacological drugs and inhibitors. Nonetheless, until now, no data are available on the effect of these inhibitors in various pathological events like atherosclerotic cardiovascular disease, neovascularization, polycystic kidney disease, and coronary artery disease.

Acknowledgments This work was supported by the Department of Higher Education, Government of Uttar Pradesh, for Centre of Excellence in Biochemistry and Indian Council of Medical Research, New Delhi. The infrastructural facilities provided by the University of Lucknow, Lucknow, are also gratefully acknowledged.

References

Abrahamson M, Mason RW, Hansson H, Buttle DJ, Grubb A, Ohlsson K (1991) Human cystatin C. role of the N-terminal segment in the inhibition of human cysteine proteinases and in its inactivation by leucocyte elastase. Biochem J 273(3):621–626

Agarwal SK (1988) Specificity of goat spleen cathepsin B. Med Sci Res 16:505–506

Agarwal SK (1990) Proteases cathepsins-A view. Biochem Edu 18(2):67–72

Agarwal SK (1991) Species dependence of mammalian cathepsin B. Med Sci Res 19:485–486

Agarwal SK, Khan MY (1987a) Does cathepsin B play a role in intracellular protein degradation? Biochem Int 15(4):785–792

Agarwal SK, Khan MY (1987b) Influence of buffer constituents on the activity of goat spleen cathepsin B. Med Sci Res 15:387–388

Agarwal SK, Khan MY (1988) A probable mechanism of inactivation by urea of goat spleen cathepsin B. Biochem J 256:609–613

Agarwal SK, Choudhury SD, Lamsal M, Khan MY (1997) Catalytic and physico-chemical characteristics of goat spleen cathepsin B. Biochem Mol Biol Int 42(6):1215–1226

Agarwal SK, Ukil A, Sharma S (2016) Purification and characterization of cathepsin B from buffalo (*Bubalus bubalis*) lung. J Biochem Technol 7(1):1051–1057

Agarwal SK, Ukil A, Singh S, Sharma S (2018) Progressive study on the physiological role and catalytic properties of buffalo lung cathepsin B. Int J Health Sci Res 8(10):51–57

Ahmad S, Agarwal SK, Khan MY (1989) Purification and some properties of buffalo spleen cathepsin B. J Biosci 14(3):261–268

Angliker H, Wikstrom P, Kirschke H, Shaw E (1989) The inactivation of the cysteinyl exopeptidase cathepsin H and C by affinity-labelling reagents. Biochem J 262:63–68
Antoniou AN, Blockwood SL, Mazzeo D, Watts C (2000) Control of Ag presentation by a single protease cleavage site. Immunity 12(4):391–398
Azaryan A, Galoyan A (1987) Human and bovine brain cathepsin L and cathepsin H: purification, physico-chemical properties, and specificity. Neurochem Res 12:207–213
Baker EN (1980) Structure of actinidin after refinement at 1.7 Å resolution. J Mol Biol 141:441–424
Baker EN, Dodson EJ (1980) Crystallographic refinement of the structure of actinidin at 1.7 Å resolution by fast fourier least-squares methods. Acta Crystallogr 36:559–572
Bansal R, Singh S, Kidwai JR (1981) Conformational studies of cathepsin B: interaction with specific antibodies. Indian J Biochem Biophys 18:110–113
Barrett AJ (1977) In: Barrett AJ (ed) Proteinases in mammalian cells and tissues. Elsevier Scientific Publishing, Amsterdam, pp 181–208
Barrett AJ, Kirschke H (1981) Cathepsin B, cathepsin H and cathepsin L. Methods Enzymol 80:535–561
Barrett AJ, McDonald JK (1980) Mammalian proteases: a glossary and bibliography-endopeptidase, vol 1. Academic, London
Barrett AJ, Kembhavi AA, Brown MA, Kirschke H, Knight CG, Tamai M, Hanada K (1982) L-trans-epoxysuccinyl-leucylamido(4-guanidino)butane (E-64) and its analogues as inhibitors of cysteine proteinases including cathepsins B, H and L. Biochem J 201:189–198
Barrett AJ, Rawlings ND, Woessner JF (1998) Handbook of proteolytic enzymes. Academic, London
Baudyš M, Meloun B, Gan-Erdene T, Pohl J, Kostka V (1990) Disulfide bridges of bovine spleen cathepsin B. Biol Chem 371:485–491
Baudys M, Meloun B, Gan-Erdene M, Fusek M, Mers M, Kostka V, Pohl J, Blake CCF (1991) S-S bridges of cathepsin B and H from bovine spleen: a basis for cathepsin B model building and possible functional implications for discrimination between exo- and endopeptidase activities among cathepsin B, H and L. Biomed Biochim Acta 50:569–577
Bohley P, Kirschke H, Langner J, Wiederanders B, Ansorge S, Hanson H (1974) In: Hanson H, Bohley P (eds) Intracellular protein catabolism. Barth, Leipzig, pp 201–209
Bond JS, Butler PE (1987) Intracellular proteases. Annu Rev Biochem 56:333–364
Bossard MJ, Tomaszek TA, Thompson SK, Amegadzie BY, Hanning CR, Jones C, Kurdyla JT, McNulty DE, Drake FH, Gowen M, Levy MA (1996) Proteolytic activity of human osteoclast cathepsin K. J Biol Chem 271(21):12517–12524
Brocklehurst K, Willenbrock F, Salih E (1987) In: Neuberger A, Brocklehurst K (eds) Hydrolytic enzymes. Elsevier, Amsterdam, pp 39–158
Bromme D, Wilson S (2011) Role of cysteine cathepsins in extracellular proteolysis. In: Parks WC, Mecham RP (eds) Extracellular matrix degradation. Springer, Berlin, pp 23–51
Brömme D, Okamoto K, Want BB, Biroc S (1996) Human cathepsin O2, a matrix protein-degrading cysteine protease expressed in osteoclasts. J Biol Chem 271:2126–2132
Bromme F, Li Z, Barnes M, Mehler E (1999) Human cathepsin V functional expression, tissue distribution, electrostatic surface potential, enzymatic characterization and chromosomal localization. Biochemistry 38:2377–2385
Buck MR, Karustis DG, Day NA, Honn KV, Sloane BF (1992) Degradation of extracellular matrix proteins by human cathepsin B from normal and tumor tissues. Biochem J 28:273–278
Bühling F, Gerber A, Häckel C, Krüger S, Köhnlein T, Brömme D, Reinhold D, Ansorge S, Welte T (1999) Expression of cathepsin K in lung epithelial cells. Am J Respir Cell Mol Biol 20 (4):612–619
Buhling F, Kouadio M, Chwieralski CE, Kern U, Hohlfeld JM, Klemm N, Friedrichs N, Roth W, Deussing JM, Peters C, Reinheckel T (2011) Gene targeting of the cysteine peptidase cathepsin H impairs lung surfactant in mice. PLoS One 6(10):1–10
Burgess JK, Boustany S, Moir LM, Weckmann M, Lau JY, Grafton K, Baraket M, Hansbro PM, Hansbro NG, Foster PS, Black JL, Oliver BG (2009) Reduction of tumstatin in asthmatic

airways contributes to angiogenesis, inflammation, and hyperresponsiveness. Am J Respir Crit Care Med 181(2):106–115
Cezari MHS, Maria L, Juliano A, Carmona AK, Juliano L (2002) Cathepsin B carboxydipeptidase specificity analysis using internally quenched fluorescent peptides. Biochem J 368(1):365–369
Chan SJ, Segundo SB, McCormick MB, Steiner DF (1986) Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. Proc Natl Acad Sci U S A 83 (11):7721–7725
Cheng XW, Shi GP, Kuzuya M, Sasaki T, Okumura K, Murohara T (2012) Role for cysteine protease cathepsins in heart disease: focus on biology and mechanisms with clinical implication. Circulation 125:1551–1562
Chilosi M, Pea M, Martignoni G, Brunelli M, Gobbo S, Poletti V, Bonetti F (2009) Cathepsin K expression in pulmonary lymphangioleiomyomatosis. Mod Pathol 22(2):161–166
Choe Y, Leonetti F, Greenbaum DC, Lecaille F, Bogyo M, Bromme D, Ellman JA, Craik CS (2006) Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J Biol Chem 281(18):12824–12832
Chornaya V, Lyannaya O (2004) Some physicochemical properties of cathepsin H from human meningioma. Exp Oncol 26:278–281
Chou PY, Fasman GD (1974) Prediction of protein conformation. Biochemistry 13(2):222–245
Choudhury SD, Lamsal M, Agarwal SK, Sharma R, Khan MY (1997) On the tissue/species dependence of cathepsin B isozymes. Mol Cell Biochem 177:89–95
Coulibaly S, Sehwihla H, Abrahamson H, Albini A, Cerni C, Clark JL, Ng KM, Katunuma N, Schlappack O, Glossl J, Mach L (1999) Modulation of invasive properties of murine squamous carcinoma cell by heterologous expression of cathepsin B and cystatin C. Int J Cancer 83:526–531
Cuervo AM, Dice JE (2000) When lysosomes get old. Exp Gerontol 35:119–131
Dean RT (1976) The roles of cathepsin B1 and D in the digestion of cytoplasmic proteins *in vitro* by lysosomal extracts. Biochem Biophys Res Commun 68:518–523
del Re EC, Shuja S, Cai J, Murnane MJ (2000) Alterations in cathepsin H activity and protein patterns in human colorectal carcinomas. Br J Cancer 82:1317–1326
Diment S, Leech MS, Stahl PD (1988) Cathepsin D is membrane-associated in macrophage endosomes. J Biol Chem 263(14):6901–6907
Dingle H (1984) Behaviour, genes, and life histories: complex adaptations in uncertain environments. In: Price PW, Slobodchikoff CH, Gaud WS (eds) A new ecology: novel approaches to interactive systems. Wiley, New York, pp 169–194
Docherty K, Carroll RJ, Steiner DF (1982) Conversion of proinsulin to insulin: involvement of a 31,500 molecular weight thiol protease. Proc Natl Acad Sci U S A 79(15):4613–4617
Docherty AJP, O'Connell J, Crabbe T, Angal S, Murphy G (1992) The matrix metalloproteinases and their natural inhibitors: prospects for treating degenerative tissue disease. Trends Biotechnol 10(6):200–207
Dodt J, Reichwein J (2003) Human cathepsin H: deletion of the mini-chain switches substrate specificity from aminopeptidase to endopeptidase. Biol Chem 384:1327–1332
Drake FH, Dodds RA, James IE, Connor JR, Debouck C, Richardson S, Lee-Rykaczewski E, Rieman D, Barthlow R, Hastings G, Gowen M (1996) Cathepsin K, but not cathepsins B, L, or S, is abundantly expressed in human osteoclasts. J Biol Chem 271:12511–12516
Drenth J, Jansonius JN, Koekoek R, Swen HM, Wolthers BG (1968) Structure of papain. Nature 218:929–932
Dufour E (1988) Sequence homologies, hydrophobic profiles and secondary structures of cathepsin B, H and L: comparison with papain and actinidin. Biochimie 70:1335–1342
Dufour E, Dive V, Toma F (1988) Delineation of chicken cathepsin L secondary structure: relationship between pH dependence activity and helix content. Biochim Biophys Acta 955 (1):58–64
Evered D, Whelan J (1978) Protein degradation in health and disease. Springer, Berlin

Faiz A, Tjin G, Harkness L, Weckmann M, Bao S, Black JL, Oliver BGG, Burgess JK (2013) The expression and activity of cathepsins D, H and K in asthmatic airways. PLoS One 8(3):e57245
Fazili KM, Qasim MA (1986) Purification and some properties of buffalo liver cathepsin B. J Biochem 100:293–299
Ferrara M, Wojcik F, Rhaissi H, Mordier S, Roux MP, Bechet D (1990) Gene structure of mouse cathepsin B. FEBS Lett 273(1–2):195–199
Fong D, Calhoun DH, Hsieh WT, Lee B, Wells RD (1986) Isolation of a cDNA clone for the human lysosomal proteinase cathepsin B. Proc Natl Acad Sci U S A 83(9):2909–2913
Fonovic M, Brömme D, Turk V, Turk B (2004) Human cathepsin F: expression in baculovirus system, characterization and inhibition by protein inhibitors. Biol Chem 385(6):505–509
Fruton JS, Irving GW, Bergmann M (1941) On the proteolytic enzymes of animal tissues. II. The composite nature of beef spleen cathepsin. J Biol Chem 138:249–262
Fuchs R, Machleidt W, Gassen HG (1988) Molecular cloning and sequencing of a cDNA coding for mature human kidney cathepsin H. Biol Chem 369:469–435
Fujishima A, Imai Y, Nomura T, Fujisawa Y, Yamamoto Y, Sugawara T (1997) The crystal structure of human cathepsin L complexed with E-64. FEBS Lett 407:47–50
Gabrijelcic D, Svetic B, Spaić D, Skrk J, Budihna M, Dolenc I, Popovic T, Cotic V, Turk V (1992) Cathepsins B, H and L in human breast carcinoma. Eur J Clin Chem Clin Biochem 30:69–74
Gairola CG, Galicki NI, Cardozo C, Lai YL, Lesser M (1989) Cigarette smoke stimulates cathepsin B activity in alveolar macrophages of rats. J Lab Clin Med 114:419–425
Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120:97–120
Gianni L, Grasselli G, Cresta S, Locatelli A, Vigano L, Minotti G (2003) Anthracyclines. Cancer Chemother Biol Response Modif 21:29–40
Gieselmann V, Pohlmann R, Hasilik A, von Figura K (1983) Biosynthesis and transport of cathepsin D in cultured human fibroblasts. J Cell Biol 97:1–5
Gocheva V, Wang HW, Gadea BB, Shree T, Hunter KE, Garfall AL, Berman T, Joyce JA (2010) IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. Genes Dev 24:241–255
Godat E, Lecaille F, Desmazes C, Duchene S, Weidauer E, Saftig P, Bromme D, Vandier C, Lalmanach G (2004) Cathepsin K: a cysteine protease with unique kinin degrading properties. Biochem J 383:501–506
Guha S, Padh H (2008) Cathepsins: fundamental effectors of endolysosomal proteolysis. Indian J Biochem Biophys 45(2):75–90
Guncar G, Podobnik M, Pungercar J, Strukelj B, Turk V, Turk D (1998) Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. Structure 6(1):51–61
Gunčar G, Klemenčič I, Turk B, Turk V, Karaoglanovic-Carmona A, Juliano L, Turk D (2000) Crystal structure of cathepsin X: a flip-flop of the ring of His23 allows carboxy-monopeptidase and carboxy-dipeptidase activity of the protease. Structure 80:305–313
Hao Y, Purtha W, Cortesio C, Rui H, Gu Y, Chen H, Sickmier EA, Manzanillo P, Huang X (2018) Crystal structures of human procathepsin H. PLoS One 13(7):1–14
Hara K, Kominami E, Katunuma K (1988) Effect of proteinase inhibitors on intracellular processing of cathepsin B, H and L in rat macrophages. FEBS Lett 231(1):229–231
Haraguchi CM, Ishido K, Kominami E, Yokota S (2003) Expression of cathepsin H in differentiating rat spermatids: immunoelectron microscopic study. Histochem Cell Biol 120:63–71
Hasnain S, Hirama T, Tam A, Mort JS (1992) Characterization of recombinant rat cathepsin B and nonglycosylated mutants expressed in yeast. J Biol Chem 267:4713–4721
Hiwasa T, Yokoyama S, Ha JM, Noguchi S, Sakiyama S (1987) c-Ha-ras gene products are potent inhibitors of cathepsins B and L. FEBS Lett 211(1):23–26

Hook G, Jacobsen JS, Grabstein K, Kindy M, Hook V (2015) Cathepsin B is a new drug target for traumatic brain injury therapeutics: evidence for E64d as a promising lead drug candidate. Front Neurol 6:178

Horn M, Maresova M, Rulisek L, Masa M, Vasiljeva O, Turk B, Erdene TG, Baudys M, Mares M (2005) Activation processing of cathepsin H impairs recognition by its propeptide. Biol Chem 386:941–947

Ii K, Hizawa K, Kominami E (1985) Different immunolocalizations of cathepsins B, H and L in the liver. J Histochem Cytochem 33:1173–1175

Illy C, Quraishi O, Wang J, Purisima E, Vernet T, Mort JS (1997) Role of the occluding loop in cathepsin B activity. J Biol Chem 272(2):1197–1202

Ishidoh K, Imajoh S, Emori Y (1987) Molecular cloning and sequencing of cDNA for rat cathepsin H: homology in propeptide regions of cysteine proteinases. FEBS Lett 226(1):33–37

Ishidoh K, Kominami E, Katunuma N, Suzuki K (1989) Gene structure of rat cathepsin H. FEBS Lett 253(1–2):103–107

Ishidoh K, Suzuki K, Katunuma N, Kominami E (1991) Gene structures of rat cathepsins H and L. Biomed Biochim Acta 50(4–6):541–547

Juhg H, Lee EY, Lee SI (1999) Age related changes in ultrastructural feature at cathepsin B and D containing neuron in rat cerebral cortex. Brain Res 844(1–2):43–54

Kalnitsky G, Chatterjee R, Singh H, Lones M, Paszkowski A (1983) Bifunctional activities and possible modes of regulation of some lysosomal cysteinyl proteases. In: Katunuma N, Umezawa H, Holzer H (eds) Proteinase inhibitors: medical and biological aspects. Springer, Berlin, pp 263–273

Kamphuis IG, Kalk KH, Swarte MB, Drenth J (1984) Structure of papain refined at 1.65 Å resolution. J Mol Biol 179:233–286

Katunuma N (2010) Posttranslational processing and modification of cathepsin and cystatins. J Signal Transduction 2010:1–8

Katunuma N, Kominami E (1986) In: Turk V (ed) Cysteine proteinases and their inhibitors. Walter de Gruyter and Co, Berlin, pp 219–227

Katunuma N, Murata E, Kakegawa H, Matsui A, Tsuzuki H, Tsuge H, Turk D, Turk V, Fukushima M, Tada Y, Asao T (1999) Structure based development of novel specific inhibitors for cathepsin L and cathepsin S *in vitro* and in vivo. FEBS Lett 458:6–10

Khan MY, Ahmad S (1987) Anomalous behavior of cathepsin B: dependence of activity and stability on salt concentration. Biochem Int 15(1):111–115

Khan MY, Ahmad S, Agarwal SK (1986) On the physiological role of mammalian cathepsin B. IRCS Med Sci 14:1141–1142

Khan MY, Agarwal SK, Ahmad S (1992) Structure-activity relationship in buffalo spleen cathepsin B. J Biochem 111:732–735

Kirschke H, Wiederanders B, Bromme D, Rinne A (1989) Cathepsin S from bovine spleen: purification, distribution, intracellular localization and action on proteins. Biochem J 264:467–473

Klemencic I, Carmona AK, Cezari MH, Juliano MA, Juliano L, Guncar G, Turk D, Krizaj I, Turk V, Turk B (2000) Biochemical characterization of human cathepsin X revealed that the enzyme is an exopeptidase, acting as carboxymonopeptidase or carboxydipeptidase. Eur J Biochem 267:5404–5412

Koga H, Mori N, Yamada H, Nishimura Y, Tokuda K, Kato K, Imoto T (1992) Endo-and aminopeptidase activities of rat cathepsin H. Chem Pharm Bull 40(4):965–970

Kominami E, Tsukahara T, Bando Y, Katunuma N (1985) Distribution of cathepsins B and H in rat tissues and peripheral blood cells. J Biochem 98:87–93

Kominami E, Tsukahara T, Hara K, Katunuma N (1988) Biosyntheses and processing of lysosomal cysteine proteinases in rat macrophages. FEBS Lett 231(1):225–228

Kornfeld S (1987) Trafficking of lysosomal enzymes. FASEB J 1(6):462–468

Kornfeld R, Kornfeld S (1980) In: Lennarz WJ (ed) The biochemistry of glycoproteins and proteoglycans. Plenum Press, New York, pp 1–34

Kos J, Schweiger A (2002) Cathepsins and cystatins in extracellular fluids useful biological markers in cancer. Radiol Oncol 36(2):176–179

Kos J, Smid A, Krasovec M, Svetic B, Lenarcic B, Vrhovec I, Skrk J, Turk V (1995) Lysosomal proteases cathepsins D, B, H, L and their inhibitors stefins A and B in head and neck cancer. Biol Chem 376:401–405

Kos J, Stabuc B, Schweiger A, Krasovec M, Cimerman N, Kopitar-Jerala N, Vrhovec I (1997) Cathepsins B, H, and L and their inhibitors stefin A and cystatin C in sera of melanoma patients. Clin Cancer Res 3:1815–1822

Lafuse WP, Brown D, Castle L, Zwilling BS (1995) IFN-gamma increases cathepsin H mRNA levels in mouse macrophages. J Leukoc Biol 57:663–609

Lah TT, Kos J (1998) Cysteine proteinases in cancer progression and their clinical relevance for prognosis. Biol Chem 379(2):125–130

Lah T, Kosorok MD, Turk V, Pain RH (1984) Conformation, structure and activation of bovine cathepsin D. Biochem J 218:601–608

Lamsal M, Agarwal SK, Choudhury SD, Khan MY (1997) Purification and tissue/species dependence of the specificity of buffalo kidney cathepsin B. Indian J Biochem Biophys 34:461–469

Lautwein A, Burster T, Lennon-Dumenil AM, Overkleeft HS, Weber E, Kalbacher H, Driessen C (2002) Inflammatory stimuli recruit cathepsin activity to late endosomal compartments in human dendritic cells. Eur J Immunol 32:3348–3357

Lecaille F, Chowdhury S, Purisima E, Brömme D, Lalmanach G (2007) The S2 subsites of cathepsins K and L and their contribution to collagen degradation. Protein Sci 16(4):662–670

Lenarčič B, Križaj I, Žunec P, Turk V (1996) Differences in specificity for the interactions of stefins A, B and D with cysteine proteinases. FEBS Lett 395(2–3):113–118

Lenney JF, Tolan JR, Sugai WJ, Lee AG (1979) Thermostable endogenous inhibitors of cathepsins B and H. Eur J Biochem 101:153–161

Li Z, Yasuda Y, Li W, Bogyo M, Katz N, Gordon RE, Fields GB, Bromme D (2004) Regulation of collagenase activities of human cathepsins by glycosaminoglycans. J Biol Chem 279:5470–5479

Liao CR, Lenney FJ (1984) Cathepsins J and K: high molecular weight cysteine proteinases from human tissues. Biochem Biophys Res Commun 124(3):909–916

Lutgens SPM, Cleutjens KBJM, Daemen MJAP, Heeneman S (2007) Cathepsin cysteine proteases in cardiovascular disease. FASEB J 21:3029–3041

Lützner N, Kalbacher H (2008) Quantifying cathepsin S activity in antigen presenting cells using a novel specific substrate. J Biol Chem 283(52):36185–36194

MacGregor RR, Hamilton JW, Kent GN, Shofstal RE, Cohn DV (1979) The degradation of proparathormone and parathormone by parathyroid and liver cathepsin B. J Biol Chem 254:4428–4433

Machleidt W, Ritonja A, PopoviE T, Kotnik M, Brzin J, Turk V, Machleidt I, Mueller-Esterl W (1986) Human cathepsins B, H and L: characterization by amino acid sequences and some kinetics of inhibition by the kininogens. In: Turk V (ed) Cysteine proteinases and their inhibitors. Walter de Gruyter and Co, Berlin/New York, pp 3–18

Mainferme F, Wattiaux R, von Figura K (1985) Synthesis, transport and processing of cathepsin D in Morris hepatoma 7777 cells and rat hepatocytes. Eur J Biochem 153:211–216

Mason RW (1989) Interaction of lysosomal cysteine proteinases with α_2-macroglobulin: conclusive evidence for the endopeptidase activities of cathepsins B and H. Arch Biochem Biophys 273 (2):367–374

Matarrese P, Ascione B, Ciarlo L, Vona R, Leonetti C, Scarsella M, Mileo AM, Catricalà C, Paggi MG, Malorni W (2010) Cathepsin B inhibition interferes with metastatic potential of human melanoma: an *in vitro* and *in vivo* study. Mol Cancer 9:207–220

Matha VL, Derocq D, Prebois C, Katunuma N, Liaudet-Coopman E (2006) Processing of human cathepsin D is independent of its catalytic function and auto-activation: involvement of cathepsins L and B. J Biochem 139:363–371

McDonald JK, Barrett AJ (1986) Mammalian proteases: a glossary and bibliography-exopeptidase, vol 2. Academic Press, London
McDonald JK, Ellis S (1975) On the substrate specificity of cathepsin B1 and B2 including a new fluorogenic substrate for cathepsin B1. Life Sci 17:1269–1276
McGrath ME, Klaus JL, Barnes MG, Bromme D (1997) Crystal structure of human cathepsin K complexed with a potent inhibitor. Nat Struct Biol 4:105–109
McGuire MJ, Lipsky PE, Thiele DL (1997) Cloning and characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C). Biochim Biophys Acta 1351:267–273
McKay MJ, Offermann MK, Barrett AJ, Bond JS (1983) Action of human liver cathepsin B on the oxidized insulin B chain. Biochem J 213:467–471
Meloun B, Baudys M, Pohl J, Pavlik M, Kostka V (1988) Amino acid sequence of bovine spleen cathepsin B. J Biol Chem 263:9087–9093
Meyer DI, Krause E, Dobberstein B (1982) Secretory protein translocation across membranes-the role of the docking proteins. Nature 297:647–650
Mort JS, Buttle DJ (1997) Cathepsin B. Int J Biochem Cell Biol 29(5):715–720
Musil D, Zucic D, Turk D, Engh RA, Mayr I, Huber R, Popovic T, Turk V, Towatari T, Katunuma N, Bode W (1991) The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. EMBO J 10(9):2321–2330
Nagler DK, Storer AC, Portaro FC, Carmona E, Juliano L, Menard R (1997) Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. Biochemistry 36(41):12608–12615
Nägler DK, Zhang R, Tam W, Sulea T, Purisima EO, Ménard R (1999) Human cathepsin X: a cysteine protease with unique carboxypeptidase activity. Biochemistry 38:12648–12654
Nishimura Y, Kato K (1987) Intracellular transport and processing of lysosomal cathepsin B. Biochem Biophys Res Commun 148(1):254–259
Ohtani O, Fukuyama K, Epstein WL (1982) Further characterization of cysteine proteinase inhibitors purified from rat and human epidermis. Biochim Biophys Acta 707(1):21–27
Ondr JK, Pham CT (2004) Characterization of murine cathepsin W and its role in cell-mediated cytotoxicity. J Biol Chem 279:27525–27533
Otto K (1971) In: Barrett AJ, Dingle JT (eds) Tissue proteinases. North Holland Publishing Co, Amsterdam, pp 181–207
Otto K, Reisenkonig H (1975) Improved purification of cathepsin Bl and B2. Biochim Biophys Acta 379:462–475
Paciucci R, Berrozpe G, Torà M, Navarro E, García de Herreros A, Real FX (1996) Isolation of tissue-type plasminogen activator, cathepsin H and non-specific cross-reacting antigen from SK-PC-1 pancreas cancer cells using subtractive hybridization. FEBS Lett 385(1–2):72–76
Polgar L, Csoma C (1987) Dissociation of ionizing groups in the binding cleft inversely controls the endo- and exopeptidase activities of cathepsin B. J Biol Chem 262(30):14448–14453
Popovic TZ, Brzin J, Kos J, Lenarcic B, Machleidt W, Ritonja A, Hanada K, Turk V (1988) A new purification procedure of human kidney cathepsin H, its properties and kinetic data. Biol Chem 369:175–183
Quinn PS, Judah JD (1978) Calcium-dependent Golgi-vesicle fusion and cathepsin B in the conversion of proalbumin into albumin in rat liver. Biochem J 172(2):301–309
Ravish I, Raghav N (2014) Curcumin as inhibitor of mammalian cathepsin B, cathepsin H, acid phosphatase and alkaline phosphatase: a correlation with pharmacological activities. Med Chem Res 23:2847–2855
Rawlings ND, Salvesen G (2013) Handbook of proteolytic enzymes, 3rd edn. Academic, London
Rawlings ND, Barrett AJ, Bateman A (2010) MEROPS: the peptidase database. Nucleic Acids Res 38:227–233
Reiser J, Adair B, Reinheckel T (2010) Specialized roles for cysteine cathepsins in health and disease. J Clin Invest 120:3421–3431
Riese RJ, Chapman HA (2000) Cathepsin and compartmentalization in antigen presentation. Curr Opin Immunol 12:107–113

Riese RJ, Mitchell RN, Villadangos JA, Shi GP, Palmer JT, Karp ER, De Sanctis GT, Ploegh HL, Chapman HA (1998) Cathepsin S activity regulates Ag presentation and immunity. J Clin Invest 101:2351–2363

Ritonja A, Popovic T, Turk V, Wiedermann K, Machleidt W (1985) Amino acid sequence of human liver cathepsin B. FEBS Lett 181:169–172

Ritonja A, Popovic T, Kotnik M, Machleidt W, Turk V (1988) Amino acid sequences of human kidney cathepsins H and L. FEBS Lett 228(2):341–345

Rossi A, Deveraux Q, Turk B, Sali A (2004) Comprehensive search for cysteine cathepsins in the human genome. Biol Chem 385(5):363–372

Ryvnyak VV, Ryvnyak EI, Tudos RV (2004) Electron histochemical localization of cathepsin L in the liver. Bull Exp Biol Med 137:90–91

Salahuddin A, Siddiqui FA, Salahuddin P (1996) Isolation, purification and properties of cathepsin B from buffalo liver. Indian J Biochem Biophys 33:292–297

Santamaria I, Pendas AM, Velasco G, Lopez-Otin C (1998) Genomic structure and chromosomal localization of the human cathepsin O gene (CTSO). Genomics 53:231–234

Santamaria I, Velasco G, Pendas AM, Paz A, Lopez-Otin C (1999) Molecular cloning and structural and functional characterization of human cathepsin F: a new cysteine proteinase of the papain family with a long propeptide domain. J Biol Chem 274:13800–13809

Schwartz WN, Barrett AJ (1980) Human cathepsin H. Biochem J 191:487–497

Scott RP, Ninjoor V, Srivastava PN (1987) Isolation and characterization of cathepsin B from rabbit testis. J Reprod Fertil 79(1):67–74

Siklos M, BenAissa M, Thatcher GR (2015) Cysteine proteases as therapeutic targets: does selectivity matter? A systematic review of calpain and cathepsin inhibitors. Acta Pharm Sin B 5(6):506–519

Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. Proc Natl Acad Sci U S A 102:11876–11881

Singh H, Kalnitsky G (1978) Separation of a new α-N-benzoylarginine-β-naphthylamide hydrolase from cathepsin B1. J Biol Chem 253:4319–4326

Singh S, Sharma S, Agarwal SK (2020) A simple purification procedure of buffalo lung cathepsin H, its properties and influence of buffer constituents on the enzyme activity. Biochem Biophys Rep 22:1–8

Sires UI, Schmid TM, Fliszar CJ, Wang ZQ, Gluck SL, Welgus HG (1995) Complete degradation of type X collagen requires the combined action of interstitial collagenase and osteoclast-derived cathepsin B. J Clin Invest 95:2089–2095

Sloane BF, Dunn JR, Honn KV (1981) Lysosomal cathepsin B: correlation with metastatic potential. Science 212(4499):1151–1153

Smith SM, Gottesman MM (1989) Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. J Biol Chem 264(34):20487–20495

Somoza JR, Zhan H, Bowman KK, Yu L, Mortara KD, Palmer JT, Clark JM, McGrath ME (2000) Crystal structure of human cathepsin V. Biochemistry 39:12543–12551

Somoza JR, Palmer JT, Ho JD (2002) The crystal structure of human cathepsin F and its implications for the development of novel immunomodulators. J Mol Biol 322:559–568

Stachowiak K, Tokmina M, Karpinska A, Sosnowska R, Wiczk W (2004) Fluorogenic peptide substrates for carboxydipeptidase activity of cathepsin B. Acta Biochim Pol 51:81–92

Starkey PM, Barrett AJ (1973) Human cathepsin B1:inhibition by α_2-macroglobulin and other serum proteins. Biochem J 131(4):823–831

Suhar A, Marks N (1979) Purification and properties of brain cathepsin B: evidence for cleavage of pituitary lipotropins. Eur J Biochem 101:23–30

Swanson AA, Martin BJ, Spicer SS (1974) Human placental cathepsin B1: isolation and some physical properties. Biochem J 137(2):223–228

Takahashi T, Schmidt PG, Tang J (1984a) Novel carbohydrate structures of cathepsin B from porcine spleen. J Biol Chem 259(10):6059–6062

Takahashi T, Dehdarani AH, Schmidt PG, Tang J (1984b) Cathepsins B and H from porcine spleen. J Biol Chem 259(15):9874–9882
Takahashi T, Yonezawa S, Dehdarani AH, Tang J (1986a) Comparative studies of two cathepsin B isozymes from porcine spleen. J Biol Chem 261(20):9368–9374
Takahashi T, Dehdarani AH, Yonezawa S, Tang J (1986b) Porcine spleen cathepsin B is an exopeptidase. J Biol Chem 261(20):9375–9381
Takahashi T, Dehdarani AH, Tang J (1988) Porcine spleen cathepsin H hydrolyzes oligopeptides solely by aminopeptidase activity. J Biol Chem 263(22):10952–10957
Takio K, Towatari T, Katunuma N, Teller DC, Titani K (1983) Homology of amino acid sequences of rat liver cathepsins B and H with that of papain. Proc Natl Acad Sci U S A 80:3666–3670
Tallan HH, Jones EJ, Fruton JS (1952) On the proteolytic enzymes of animal tissues: beef spleen cathepsin C. J Biol Chem 194:793–705
Tan Y, Osatomi K, Hara K (2006) Gene structure of carp *Cyprinus carpio* cathepsin B. Fish Sci 72:673–678
Taniguchi T, Mizuochi T, Towatari T (1985) Structural studies on the carbohydrate moieties of rat liver cathepsins B and H. J Biochem 97(3):973–976
Tchoupé JR, Moreau T, Gauthier F, Bieth JG (1991) Photometric or fluorometric assay of cathepsin B, L and H and papain using substrates with an aminotrifluoromethylcoumarin leaving group. Biochim Biophys Acta Protein Struct Mol Enzymol 1076(1):149–151
Tezuka K, Tezuka Y, Maejima A, Sato T, Nemoto K, Kamioka H, Hakeda Y, Kumegawa M (1994) Molecular cloning of a possible cysteine proteinase predominantly expressed in osteoclasts. J Biol Chem 269(2):1106–1109
Tolosa E, Li W, Yasuda Y, Wienhold W, Denzin LK, Lautwein A, Driessen C, Schnorrer P, Weber E, Stevanovic S, Kurek R, Melms A, Bromme D (2003) Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. J Clin Invest 112:517–526
Towatari T, Katunuma N (1978) Crystallization and amino acid composition of cathepsin B from rat liver lysosomes. Biochem Biophys Res Commun 83(2):513–520
Tsuge H, Nishimura T, Tada Y, Asao T, Turk D, Turk V, Katunuma N (1999) Inhibition mechanism of cathepsin L-specific inhibitors based on the crystal structure of papain–CLIK148 complex. Biochem Biophys Res Commun 266(2):411–416
Tsushima H, Ueki A, Matsuoka Y, Mihara H, Hopsu-Havu VK (1991) Characterization of a cathepsin-H-like enzyme from a human melanoma cell line. Int J Cancer 48:726–732
Turk B, Turk D, Turk V (2000) Lysosomal cysteine proteases: more than scavengers. Biochim Biophys Acta 1147:98–111
Turk V, Turk B, Turk D (2001) Lysosomal cysteine proteases: facts and opportunities. EMBO J 20 (17):4629–4633
Turk D, Turk B, Turk V (2003) Papain-like lysosomal cysteine proteases and their inhibitors: drug discovery targets? Biochem Soc Symp 70:15–30
Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, Turk D (2012) Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochim Biophys Acta 1824:68–88
Turkenburg JP, Lamers MBAC, Brzozowski AM, Wright LM, Hubbard RE, Sturt SL, Williams DH (2002) Structure of a Cys25 → Ser mutant of human cathepsin S. Acta Crystallogr D Biol Crystallogr 58:451–445
Uchiyama Y, Wagmi S, Sato N, Watanabe T, Ishido K, Kominami E (1994) Cell and tissue distribution of lysosomal cysteine proteinases, cathepsins B, H, and L, and their biological roles. Acta Histochem Cytochem 27:351–372
Vasiljeva O, Dolinar M, Turk V, Turk B (2003) Recombinant human cathepsin H lacking mini chain is an endopeptidase. Biochemistry 42:13522–13528
Vasiljeva O, Papazoglou A, Kruger A, Brodoefel H, Korovin M, Deussing J, Augustin N, Nielsen BS, Almholt K, Bogyo M, Peters C, Reinheckel T (2006) Tumor cell-derived and macrophage-

derived cathepsin B promotes progression and lung metastasis of mammary cancer. Cancer Res 66:5242–5250
Verma S, Dixit R, Pandey KC (2016) Cysteine proteases: modes of activation and future prospects as pharmacological targets. Front Pharmacol 7:107
Victor BC, Anbalagan A, Mohamed MM, Sloane BF, Cavallo-Medved D (2011) Inhibition of cathepsin B activity attenuates extracellular matrix degradation and inflammatory breast cancer invasion. Breast Cancer Res 13(6):R115
Vidak E, Javoršek U, Vizovišek M, Turk B (2019) Cysteine cathepsins and their extracellular roles: shaping the microenvironment. Cells 8:264–288
Von Figura K, Hasilik A (1986) Lysosomal enzymes and their receptors. Annu Rev Biochem 55:167–193
Waghray A, Keppler D, Sloane BF, Schuger L, Chen YQ (2002) Analysis of a truncated form of cathepsin H in human prostate tumor cells. J Biol Chem 277:11533–11538
Walter P, Blobel G (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature 299:691–698
Wang B, Shi GP, Yao PM, Li Z, Chapman HA, Bromme D (1998) Human cathepsin F: molecular cloning, functional expression, tissue localization, and enzymatic characterization. J Biol Chem 273:32000–32008
Watanabe M, Watanabe T, Ishii Y, Matsuba H, Kimuna S, Fujita T, Kominami E, Katunuma N, Uchiyama Y (1988) Immunocytochemical localization of cathepsins B, H and their endogenous inhibitor, cystatin β in islet endocrine cells of the rat pancreas. J Histochem Cytochem 36:783–791
Weiss RB, Donehower RC, Wiernik PH, Ohnuma T, Gralla RJ, Trump DL, Baker JR Jr, Van Echo DA, Von Hoff DD, Leyland-Jones B (1990) Hypersensitivity reactions from taxol. J Clin Oncol 8(7):1263–1268
Wex T, Levy B, Smeekens SP, Ansorge S, Desnick RJ, Bromme D (1998) Genomic structure, chromosomal localization, and expression of human cathepsin W. Biochem Biophys Res Commun 248:255–261
Willenbrock F, Brocklehurst K (1985) Preparation of cathepsin B and H by covalent chromatography and characterization of their catalytic sites by reaction with a thiol- specific two-protonic-state reactivity probe. Biochem J 227:511–519
Willstätter R, Bamann E (1929) Über die proteasen der magenschleimhaut. Erste abhandlung über die enzyme der leukocyten. Hoppe Seylers Z Physiol Chem 180:127–143
Wolters PJ, Chapman HA (2000) Importance of lysosomal cysteine proteases in lung disease. Respir Res 1:170–177
Wu SM, Huang YH, Yeh CT, Tsai MM, Liao CH, Cheng WL, Chen WJ, Lin KH (2011) Cathepsin H regulated by the thyroid hormone receptors associate with tumor invasion in human hepatoma cells. Oncogene 30:2057–2069
Wu H, Du Q, Dai Q, Ge J, Cheng X (2018) Cysteine protease Cathepsins in atherosclerotic cardiovascular diseases. J Atheroscler Thromb 25(2):111–123
Yamamoto K, Kamata O, Kato Y (1984) Separation and properties of three forms of cathepsin H-like cysteine proteinase from rat spleen. J Biochem 95:477–484
Yasuda Y, Li Z, Greenbaum D, Bogyo M, Weber E, Bromme D (2004) Cathepsin V, a novel and potent elastolytic activity expressed in activated macrophages. J Biol Chem 279 (35):36761–36770
Yoshida A, Ohta M, Kuwahara K, Cao MJ, Hara K, Osatomi K (2015) Purification and characterization of cathepsin B from the muscle of horse *Mackerel Trachurus japonicas*. Mar Drugs 13 (11):6550–6565
Zavasnik-Bergant T, Turk B (2007) Cysteine proteases: destruction ability versus immunomodulation capacity in immune cells. Biol Chem 388(11):1141–1149
Zhong YJ, Shao LH, Li Y (2013) Cathepsin B-cleavable doxorubicin prodrugs for targeted cancer therapy. Int J Oncol 42(2):373–383

Zvonar-Popovic T, Kregar I, Turk V (1979) Isolation of cathepsin B and α-N-benzoylarginine-β-naphtylamide hydrolase by covalent chromatography on activated thiol Sepharose. Croat Chem Acta 52:411–416

Zvonar-Popovic T, Lah T, Kregar I, Turk V (1980) Some characteristics of cathepsin B and α-N-Benzoylarginine-β-Naphthylamide hydrolase from bovine lymph nodes. Croat Chem Acta 53 (3):509–517

17 An Insight into the Importance of Ferritins in the Physiology of *Mycobacterium tuberculosis*: Unique Structural and Functional Properties

Garima Khare, Prachi Nangpal, and Anil Kumar Tyagi

Abstract

Iron is an essential element required by most of the living organisms and acts as a cofactor for many enzymes involved in various essential cellular processes such as respiration and DNA replication. As much as it is crucial for performing major cellular functions, its excess can be detrimental to the cell by its participation in Fenton reaction, which results in the production of harmful hydroxyl radicals. Hence, iron homeostasis is a vital part of cellular physiology, which is tightly regulated by various genes involved in iron acquisition and storage. Ferritins belong to the major superfamily of iron storage proteins having a spherical macromolecular structure with a cage-like cavity and play a pivotal role in the maintenance of iron homeostasis. These proteins play a dual role by acting as a source of iron under conditions of iron scarcity as well as serving as iron quenchers under excess iron conditions. The family of ferritin proteins comprises of three subtypes, namely ferritin (Ftn) present across both eukaryotes and prokaryotes, heme-bound bacterioferritin (Bfr), and DNA-binding protein from starved cells (Dps) found only in prokaryotes. *Mycobacterium tuberculosis*, one of the most deadly pathogens, is responsible for killing millions of humans, possesses both kinds of iron-storing proteins-ferritin (BfrB) as well as bacterioferritin (BfrA), both of which are required for the pathogenesis of this deadly bacteria. Besides, the presence of both kinds of ferritin-like molecules in the pathogen is linked to distinct functions they perform in *M. tuberculosis* physiology. Moreover, structural properties of BfrB, including certain key residues at its threefold and fourfold channels, are considered as interface hot-spot residues required for BfrB oligomerization and assembly formation. Additionally, unlike other ferritins, BfrB of *M. tuberculosis* possesses an extended C-terminus region, which is implicated in playing a role in providing thermal stability to the protein.

G. Khare (✉) · P. Nangpal · A. K. Tyagi
Department of Biochemistry, University of Delhi South Campus, New Delhi, India

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_17

Keywords

Bacterioferritins · Ferritins · Fenton's reaction · Three-fold and four-fold channel · Detoxification · Iron homeostasis

17.1 Introduction

Iron is an essential metal for most of the organisms living on this planet. Its participation in the key cellular processes, including DNA replication, respiration, electron transfer chain, and redox reactions, makes this element an extremely crucial part of cellular physiology. However, iron levels in the cell need to be tightly regulated as its excess causes deleterious effect by undergoing Fenton reaction with the by-products of oxidation, resulting in the formation of toxic superoxide radicals, which in turns leads to DNA and protein damage. Moreover, for most of the organisms including important pathogens like *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa* and *Salmonella typhimurium*, iron is indispensable for survival, and hence, iron scarcity also leads to cell death (Meyer et al. 1996; Bullen et al. 1974; Lounis et al. 2001; Rodriguez and Smith 2003, Sritharan 2016; Furman et al. 1994). Therefore, dedicated machinery has been devised by various bacteria comprising of multiple proteins to carry out iron acquisition and storage that interplay to maintain iron homeostasis.

Iron acquisition/uptake is a vital process to cater to the conditions of iron scarcity, which can otherwise lead to cell death. In general, animals acquire iron from dietary sources, while plants and microorganisms have various strategies for scavenging iron. One of the primary mechanisms by which microorganisms acquire iron is via diffusion across the cellular membrane (Górska et al. 2014; Sheldon and Heinrichs 2015). Although there is a predominance of iron in the environment, this mechanism is not a preferred one due to the insoluble nature of this metal, which makes the bioavailability of iron very low. Hence, as a result, microorganisms must utilize other means of sequestering iron. These include several heme-uptake pathways present at the cellular surface and numerous iron-scavenging high-affinity siderophores (Sheldon and Heinrichs 2015). These siderophores are an incredibly diverse class of biomolecules that mainly fall into three categories: (1) hydroxamates, (2) catecholates, and (3) carboxylates. A tug of war ensues between the host and the pathogen where the host tries to restrict the levels of available iron to effectively make the pathogen iron-deficient (a situation known as nutritional immunity) and the pathogen, in turn, secretes high-affinity siderophores to sequester iron from the host and circumvent iron paucity (Sheldon and Heinrichs 2015; Golonka et al. 2019; Weinberg 1975). Thus, for microbial pathogens including *Mycobacterium tuberculosis*, *Yersinia pestis*, *Legionella*, and *Staphylococcus aureus*, the acquisition of iron is extremely important to survive the iron-deficient environment of the host (Miethke and Marahiel 2007; Reddy et al. 2013; Sebbane et al. 2010; Fetherston et al. 2010; Cianciotto 2015; Hammer and Skaar 2011).

As much as it is crucial to acquire iron for carrying out various cellular processes, it is imperative to keep its levels under control as a high level of iron is capable of inflicting damage to the cells. The versatile catalytic nature of iron allows it to participate in single-electron transfers by interconverting between the Fe^{3+} (ferric) and Fe^{2+} (ferrous) oxidation-redox states. Ferrous iron is dangerous as it can undergo Fenton reaction and cause injury to the cell, while ferric iron is more stable but insoluble under physiological conditions. Hence, under conditions of iron excess, the cell has devised iron storage proteins belonging to the ferritin family that can store iron by converting ferrous to ferric mineral by using their ferroxidase activity (Ebrahimi et al. 2015). Ferritins are almost ubiquitously present, and apart from performing their primary function of storing iron, they are also involved in carrying out other diverse functions like preventing the cells from oxidative stress and hypoxic stress (Ebrahimi et al. 2015; Honarmand Ebrahimi et al. 2015; Arosio et al. 2017).

This chapter provides a comprehensive review of the importance of iron homeostasis in view of the iron storage proteins, majorly focusing on their importance, structure, and physiological role in the context of *Mycobacterium tuberculosis*, one of the most damaging pathogens of the world.

17.2 Classification of Ferritins: Insight into Their Function

The history of ferritins dates back to 1937 when these molecules were first crystallized and given its nomenclature (Laufberger 1937). They are the major class of proteins that act as an iron reservoir and have a capacity of withholding up to 4000–5000 Fe atoms in the mineral form (Arosio et al. 2017; Laufberger 1937). Ferritins are globular proteins comprising of 24 subunits that combine to form an iron core cavity in its center (Fig. 17.1).

Each subunit of the ferritin protein comprises of a characteristic four α-helical bundle (helices A–D) and an additional loop linking helix B to helix C. All the 24 subunits arrange themselves in an octahedral 4-3-2 symmetry to form the quaternary structure enclosing the hollow iron cavity (Fig. 17.2) (Ebrahimi et al. 2015; Honarmand Ebrahimi et al. 2015; Arosio et al. 2017; Khare et al. 2011; Harrison and Arosio 1996). Although the primary sequences of various homologous ferritins vary, there is a striking similarity in their tertiary and quaternary structures.

In eukaryotes, the ferritin protein is composed of heteropolymers of three different type of subunits based on their molecular weights: L ("light," 20 kDa) chain, M ("middle," 21 kDa) chain, and H ("heavy," 22.8 kDa) chain (Honarmand Ebrahimi et al. 2015; Arosio et al. 1978). Among these different types of subunits, only H-chain and M-chain can convert the ferrous iron to ferric iron (Honarmand Ebrahimi et al. 2015). On the other hand, in bacteria and archaea, ferritin molecules are made up of identical subunits (~20 kDa) with each component capable of executing the ferroxidase activity (Harrison and Arosio 1996). Each subunit of the ferritin macromolecule has iron-binding sites (Fig. 17.3). Other important members of this ferritin superfamily are the bacterioferritins (Bfrs) composed of

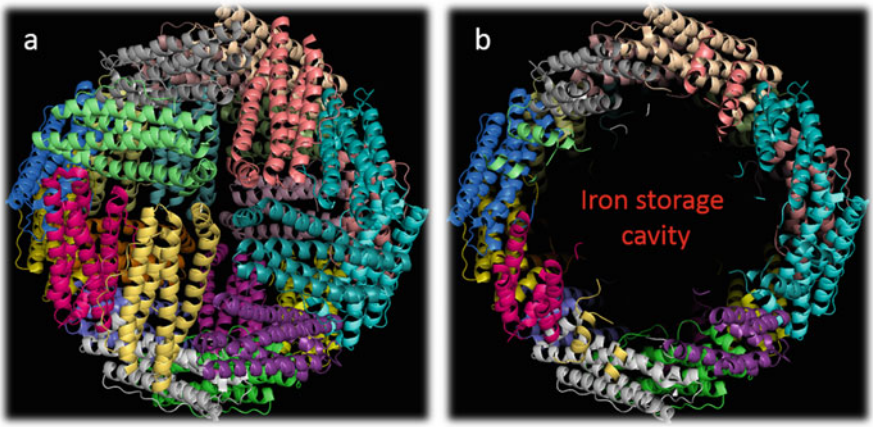


Fig. 17.1 Structure of *M. tuberculosis* ferritin (BfrB). (**a**) The ferritin subunits assemble into a spherical macromolecular assembly of 24 subunits (**b**) The globular ferritin proteins have an iron storage cavity in their center

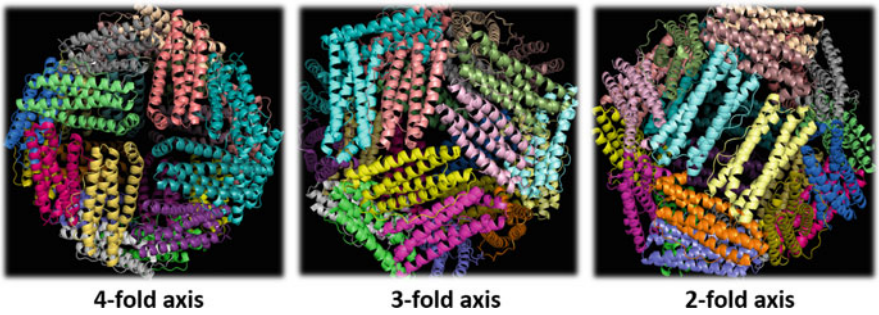


Fig. 17.2 Quaternary structural arrangement of *M. tuberculosis* ferritin (BfrB). The 24 subunits arrange themselves into four-, three-, and twofold symmetrical axis, generating an octahedral 432 symmetry in these proteins

24 homologous subunits that are found in bacteria. The only difference between ferritins and Bfrs is the ability of Bfrs to bind the heme molecule at the interface of its two individual subunits (Fig. 17.3) (Carrondo 2003). Hence, each Bfr molecule contains ~12 heme moieties (Carrondo 2003). DNA-binding protein from starved cells (Dps) and Dps-like proteins which are found only in the prokaryotes also belong to this ferritin superfamily which is composed of 12 subunits that assemble to form a spherical structure in 23 symmetrical tetrahedral arrangement (Haikarainen and Papageorgiou 2010; Calhoun and Kwon 2011).

Iron mineralization into the hollow cavity involves a cascade of complex reactions. Each subunit of ferritin and Bfr proteins contains the ferroxidase activity, which utilizes cellular oxidants like molecular oxygen, hydrogen peroxide to oxidize the ferrous form to ferric iron and store iron in its mineral form (ferrihydrite) in the

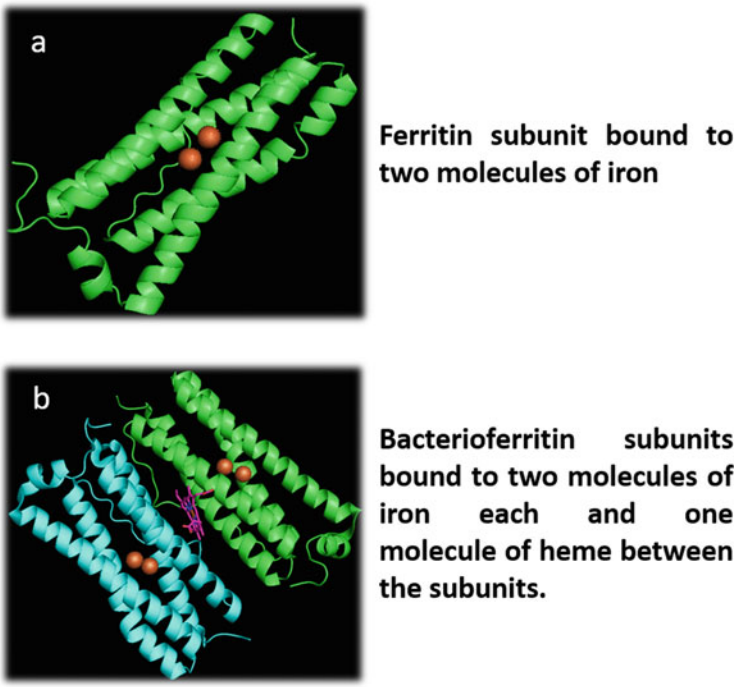


Fig. 17.3 Iron and heme-binding sites in ferritins/Bfrs, (**a**) The iron-binding sites are present between each of the ferritin subunits of *M. tuberculosis* BfrB. (**b**) The heme-binding site is present in between two subunits facing each other at the twofold symmetrical axis in *M. tuberculosis* BfrA

hollow cavity representing an iron reservoir for the cell (Khare et al. 2011; Bou-Abdallah 2010; Levi et al. 1988). The first step in iron storage is iron-binding, followed by the migration of ferrous ions to the catalytic ferroxidase site (Bou-Abdallah 2010). This is mediated through various channels and pores present at the four-, three-, and a twofold axis of the protein (Tosha et al. 2010). These channels and pores are also responsible for the influx and efflux of other molecules, including oxidants, reductants, and chelators (Khare et al. 2011; Tosha et al. 2010). The ferroxidase center catalyzes the conversion of ferrous ions to ferric ions, which are then sequestered in the iron storage cavity of the protein. While ferritins and Bfrs can accommodate ~4000 iron molecules, Dps and Dps-like proteins, are smaller in size, have lower iron storage capacity and hence are majorly involved in preventing DNA from oxidative damage (Calhoun and Kwon 2011; Chiancone and Ceci 2010). Moreover, ferritins and Bfrs have their ferroxidase center located within each subunit, while Dps and Dps-like proteins possess intersubunit ferroxidase catalytic activity (Khare et al. 2011).

17.3 The Interplay of Iron Acquisition and Storage Proteins: Highlighting *M. tuberculosis* Pathogenesis

M. tuberculosis, the causative agent of tuberculosis, is one of the most successful pathogens haunting humanity since time immemorial. Its existence has taken a massive toll on human lives, accounting for 1.2 million deaths among HIV-negative people and an additional 0.251 million deaths among HIV-positive population in the year 2018 (WHO 2019).

M. tuberculosis majorly infects humans through the aerosol route and lodges itself in the phagosomes of alveolar macrophages. Similar to many other organisms, *M. tuberculosis* also utilizes iron as a cofactor for many enzymes to carry out important and essential reactions required by the cell for its survival (Sauton 1912). In addition, the pathogen cannot survive in the absence of iron. Hence, iron homeostasis plays a crucial part in *M. tuberculosis* virulence. Moreover, due to an essential requirement of iron by the pathogen, it competes with the host for bringing the host iron to itself (Sritharan 2016). Upregulation of transferrins, lactoferrins, and ferritins by the human body results in iron-limiting conditions in response to bacterial infection (Kochan 1973; Lounis et al. 2001; Muller et al. 1983; Weinberg 1984). Moreover, IFN-γ activation of *M. tuberculosis*-infected macrophages was shown to keep the iron levels extremely low in the phagosomes by upregulating natural resistance-associated macrophage protein (Nramp1) (Searle et al. 1998; Cellier et al. 2007). Thus, in a mammalian host, free iron availability is far lesser (estimated at 10^{-18} M) than that required for bacterial survival (Neilands 1995; Weinberg 1999; Bullen 1981). Several reports also provide evidence that *M. tuberculosis* is exposed to low iron conditions during growth in human macrophages and lungs (Timm et al. 2003; Schnappinger et al. 2003). To overcome this stressful condition by the host, *M. tuberculosis* has devised strategies to scavenge iron with the help of mycobactins (MBTs), which are secretory siderophores (Ratledge 1999; Wells et al. 2013). A 20-fold increase in the biologically available iron has been observed with the help of MBTs within the phagosomal compartments (Wells et al. 2013; Wagner et al. 2005). There are two kinds of siderophores, namely MBTs and exochelins, which differ in the presence or absence of the 2-hydroxyphenyloxazoline-ring, respectively (De Voss et al. 1999). *M. tuberculosis* only produces MBTs, which are derived from salicylic acid, whereas saprophytic mycobacteria like *M. smegmatis* produces both kinds of siderophores (De Voss et al. 2000). These MBTs have a very high affinity toward iron, which aid in sequestrating iron from the iron-limiting host environment (Wagner et al. 2005). *M. tuberculosis* secretes two types of structurally different MBTs, namely a less polar cell-wall-associated form of MBT and a more polar secretory form, carboxymycobactin (CMBT) (Quadri et al. 1998). These two kinds of MBTs differ in their acyl group, which gives them different physical properties (De Voss et al. 2000; Quadri et al. 1998). It is currently unclear why *M. tuberculosis* synthesizes two kinds of MBTs; however, it is proposed that the more soluble form of MBTs binds to free iron available in the environment, whereas the membrane-bound form of MBTs is thought to act as an ionophore to shuttle the iron inside from the extracellular

milieu (Gobin and Horwitz 1996; Golden et al. 1974). Quadri et al. identified the genes that are involved in the MBTs' core structure biosynthesis (Quadri et al. 1998). These genes are assembled as part of a ten-gene cluster known as *mbt* cluster, designated as *mbt-1* or *mbtA-J* (Quadri et al. 1998). Further, the *mbt-2* cluster comprising of four genes *mbtK* to *mbtN* is responsible for the addition of a hydrophobic aliphatic side chain on the core structure of MBTs (Krithika et al. 2006). Several reports have shown the importance of genes involved in the synthesis of siderophores in bacterial survival as well as pathogenesis (Reddy et al. 2013; De Voss et al. 2000). For instance, *mbtB* mutant of *M. tuberculosis* was found to be defective in the synthesis of MBTs and showed impaired growth in THP-1 macrophages (De Voss et al. 2000). Reddy et al., in their seminal article, showed that inactivation of the *mbtE* gene of *M. tuberculosis* renders the pathogen unable to synthesize siderophores (Reddy et al. 2013). Moreover, the cellular morphology of the *mbtE* mutant was shown to be altered. The mutant exhibited attenuated growth in liquid culture as well as in macrophages. This altered phenotype and impairment in the growth of the pathogen were complemented by supplementing the growth medium with exogenous MBTs (Reddy et al. 2013). In addition, guinea pigs infected with the *mbtE* mutant showed significantly reduced bacillary load as well as pathological damage in the various organs of the infected animals (Reddy et al. 2013). These studies implicated the essentiality of the MBTs biosynthesis pathway in the survival of *M. tuberculosis* under low iron conditions of the host and are therefore an attractive target for drug discovery.

Although much has been learned about the synthesis of MBTs, very few studies have identified and shown the importance of genes involved in iron uptake. Rodriguez et al. identified two genes (*irtA* and *irtB*), which encode for an ABC transporter, responsible for efficient transportation of iron molecules by using Fe-carboxymycobactin as the source (Rodriguez and Smith 2006). Moreover, *irtA* and *irtB* were shown to play an important role in the growth of *M. tuberculosis* in human macrophages and mouse lungs, thereby suggesting the in vivo significance of this transporter complex (Rodriguez and Smith 2006). In another study published by Siegrist et al., it was shown that mycobacterial mutants of BCG and *M. smegmatis* lacking the Esx-3 system was defective in acquiring iron, impaired in their growth in vitro and in macrophages, despite the fact that they were able to synthesize MBTs, implicating the role of Esx-3 system in iron import (Siegrist et al. 2009). In a genome-wide screen of *M. tuberculosis* transposon mutagenesis library, it was observed that transposon mutants in *esx-3* locus could not be recovered, suggesting the essentiality of this locus for *M. tuberculosis* survival (Sassetti et al. 2003). Two iron-regulated genes, *mmpS4 a*nd *mmpS5*, which encode for membrane proteins, were found to be important for *M. tuberculosis* growth under low iron conditions (Wells et al. 2013). A double mutant of *mmpS4* and *mmpS5* showed reduced synthesis as well as secretion of MBTs. The double mutant was shown to be attenuated in mice model of tuberculosis, suggesting the involvement of these membrane-associated proteins in the survival and virulence of *M. tuberculosis* in the host. Protein pull-down assay showed the association of MmpS4 with transporter protein MmpL4 through the periplasmic loop of the latter, whereas MmpS5 was

demonstrated to be associated with MmpL5 (Wells et al. 2013). Based on the data, MmpS4/MmpL4 and MmpS5/MmpL5 were identified to be part of the siderophore export system (Wells et al. 2013).

Furthermore, high amounts of iron in humans and mice were shown to exacerbate the disease (Gangaidzo et al. 2001; Kochan 1969). Hence, it is also important for the pathogen to regulate its intracellular iron levels and maintain iron homeostasis. In general, the regulation of genes involved in the acquisition, transportation, and storage of iron is under tight control in bacterial systems. These transcriptional regulators are divided into subcategories: Fur and DtxR family of transcriptional factors (Doukhan et al. 1995). Ferric uptake regulator (Fur) was identified in *Escherichia coli*, and its homologs are found in many prokaryotes, including gram-positive as well as gram-negative bacteria (Rodriguez et al. 2002; Hantke 1981, 2001; Deng et al. 2015). Fur transcriptional regulator in *E. coli* was shown to control the expression of more than 60 genes involved in various pleiotropic reactions ranging from iron homeostasis, toxin metabolism, and oxidative stress (Hantke 2001; Touati 2000). In *P. aeruginosa*, the master regulator for iron uptake is Fur that controls the expression of genes involved in siderophore synthesis, which is required for the survival of the pathogen (Minandri et al. 2016; Ochsner et al. 2002; Cornelis et al. 2009). Iron acquisition is controlled by the Fur family of transcriptional regulators in *S. aureus* and is essential for colonization and subsequent pathogenesis (Friedman et al. 2006; Hammer and Skaar 2011). Fur-deficient mutant of *S. aureus* was shown to exhibit the phenotype of an iron-starved organism (Hammer and Skaar 2011). Further, DtxR (iron-dependent toxin regulator) was identified in *Cornybacterium diphtheria* and later found to be present in actinomycetes, including mycobacteria and streptomycetes (Doukhan et al. 1995; Boyd et al. 1990; Schmitt and Holmes 1991). There are four major iron-responsive transcriptional factors found in *M. tuberculosis*; FurA and FurB that complex to form FuR transcription factor and IdeR and SirR that belong to the DtxR family of regulators (Schmitt et al. 1995; Günter-Seeboth and Schupp 1995). Among these four, IdeR has been extensively studied for its structure as well as function (Chou et al. 2004; Rodriguez et al. 2002). Iron-dependent regulator (IdeR) was first shown to negatively regulate the siderophore synthesis in *M. smegmatis* (Dussurget et al. 1996). Gold et al. identified the genes regulated by IdeR in *M. tuberculosis* (Gold et al. 2001). IdeR exhibits close homology to DtxR of *C. diphtheria*, both structurally and functionally, sharing 92% identity in its DNA-binding domain and 100% similarity in the amino acid sequence (Gold et al. 2001). Moreover, IdeR was able to complement the mutant of *dtxR* for iron-dependent control of genes regulated by DtxR in *C. Diphtheria* (Schmitt et al. 1995). Hence, based on the sequences of *dtxR* boxes, Gold et al. identified IdeR boxes in the genome of *M. tuberculosis* (Gold et al. 2001). About 40 genes were belonging to several categories that were identified to contain the putative IdeR boxes (Gold et al. 2001). In the category of iron metabolism genes, iron acquisition genes (*mbtA–mbtB*, *mbtI*, *rv1348*, *rv1347c*) and iron storage genes (*bfrA, bfrB*) were identified to contain the IdeR binding boxes (Gold et al. 2001). IdeR is considered a dual transcriptional regulator, which functions as a repressor for MBT synthesis genes and also an activator for iron storage genes. IdeR

senses the levels of intracellular iron and accordingly regulates the transcriptional machinery in favor of either MBT synthesis or iron storage genes synthesis (Gold et al. 2001). Thus, under conditions of iron availability, expression of *mbt1/mbt2* locus is repressed, whereas the transcription of the iron storage genes *bfrA* and *bfrB* is activated (Gold et al. 2001). IdeR has been demonstrated to be essential for *M. tuberculosis*, as highlighted by the fact that its disruption is only possible in saprophytic *Mycobacterium* such as *M. smegmatis*, whereas, a deletion mutant of *ideR* in *M. tuberculosis* is lethal (Rodriguez et al. 2002). Hence, only a conditional mutant of *ideR* in *M. tuberculosis* could be generated. Pandey et al. demonstrated that conditional mutant of *ideR* exhibited iron unresponsiveness with ~100 times more synthesis of MBTs in comparison to the parental strain (Pandey and Rodriguez 2014). There was a constitutive expression of the siderophore biosynthesis genes observed in the *ideR* conditional mutant, which was linked to the increase in the synthesis of MBTs (Pandey and Rodriguez 2014). Measurement of intracellular iron revealed the accumulation of toxic iron levels within the *ideR* conditional knockout (Pandey and Rodriguez 2014). Moreover, the high concentration of intracellular iron also exacerbated the Fenton reaction and subsequent oxidative stress within the IdeR-deficient cells (Pandey and Rodriguez 2014). Besides, a conditional mutant of *ideR* was unable to replicate in macrophages as well as in mice highlighting that IdeR is indispensable for *M. tuberculosis* survival (Pandey and Rodriguez 2014).

Furthermore, apart from regulating the expression of iron acquisition genes, IdeR also controls the expression of iron storage genes *bfrA* and *bfrB* in response to varying iron concentrations (Pandey and Rodriguez 2014; Gold et al. 2001). Functional and structural characterization of IdeR revealed that it binds to iron at the metal-binding site, which leads to its activation by dimer formation (Chou et al. 2004). This activation brings about a conformational change, which orients the DNA-binding helix of each monomer for binding to the DNA (Chou et al. 2004). Interestingly, *bfrA* gene expression is regulated by three different promoters, namely P_{low1}, P_{low2}, and P_{high}. P_{low1} and P_{low2} are located upstream of $P_{high,}$ and the IdeR box is located at the promoter region of P_{high}. Under conditions of iron starvation, IdeR is not bound to iron and hence is unable to bind to the IdeR box, which is located upstream of P_{high} (Gold et al. 2001). As a result, RNA polymerase can transcribe *bfrA* mRNA from both P_{low} and P_{high} (Gold et al. 2001). On the contrary, under conditions of iron excess, IdeR binds at the IdeR box and blocks transcription from P_{low} promoter, whereas acts as an activator for expression via P_{high} (Gold et al. 2001). Thus, it appears that BfrA would be playing a crucial role in iron homeostasis. There is an IdeR box found upstream of the promoter region of the *bfrB* gene that is suggested to be involved in IdeR-mediated positive regulation of *bfrB* gene in the presence of high iron (Gold et al. 2001).

Owing to their importance in the physiology of the pathogen, these iron storage proteins are crucial for the survival of many pathogens. The next section describes the importance of these proteins for *M. tuberculosis* physiology as well as pathogenesis.

17.4 Ferritins and Bfrs: Their Importance and Role in the Physiology of *M. tuberculosis*

17.4.1 BfrA and BfrB Impart Protection from the Oxidative Damage

Iron storage proteins are essential for the physiology as well as the survival of *M. tuberculosis* and other pathogens. Moreover, the ability of these proteins to scavenge the intracellular iron also helps in preventing the Fenton reaction and subsequent oxidative damage. Hence, these proteins not only act as an iron reservoir for the cell but also prevent the metal from mediating iron toxicity. The *katA* gene of *P. aeruginosa*, which plays a role in preventing the oxidative damage, is located adjacent to the *bfrA* gene. The *bfrA* mutant exhibited a decrease in its catalase activity as a result of reduced KatA synthesis in *P. aeruginosa*, which was entirely restored by the complementation with the wild-type copy of the *bfrA* gene. This observation suggested that the iron stored in BfrA is involved in the formation of the cofactor heme of KatA (Ma et al. 1999). In another study, *ftnA* and *bfr* mutants of *E. coli* exhibited impaired growth in iron-deficient media, after being cultured in iron-rich conditions (Abdul-Tehrani et al. 1999). Moreover, *E. coli* carrying mutations in *ftnA*, *bfr*, and *fur* exhibited enhanced sensitivity to hydroperoxides, as a result of an increase in the production of reactive ferrous iron (Abdul-Tehrani et al. 1999). In a study by Reddy et al., *M. tuberculosis* gene deletion mutants of *bfrA* (Rv1876) and *bfrB* (Rv3841) were generated (Reddy et al. 2012). Mycobacteria are known to be continuously exposed to various stresses such as phagosomal acidification, nitrosative stress, and oxidative-redox stress generated by the activated macrophages, the primary niche of *M. tuberculosis* (Boelaert et al. 2007). Single mutants of *M. tuberculosis* lacking either BfrA or BfrB showed marginal resistance to hydrogen peroxide and plumbagin based on zone inhibition assay. However, the deletion of both the genes in *M. tuberculosis* (M.tbΔ*bfrAbfrB*) was found to enhance the susceptibility of the double mutant strain to oxidative stress by carrying out zone inhibition assay (Reddy et al. 2012). Further, their relative roles in protecting against oxidative stress under varying iron conditions were assessed. In general, the iron molecules from iron cluster proteins can be reductively leached out by the oxygen radicals produced in the cell through various metabolic processes (Keyer and Imlay 1996). The released iron molecules in the presence of hydrogen peroxide lead to a Fenton reaction that can result in the production of toxic molecules such as ferric ions and hydroxyl radicals (Touati 2000). Hence, pathogens utilize these iron storage proteins to avoid the damage caused by the leaching of released iron under oxidative stress. Therefore, Khare et al. performed the growth kinetics of these single as well as double mutants of *M. tuberculosis* cultured under low and high iron conditions. In the presence of hydrogen peroxide, single mutants also displayed growth defective phenotype in comparison to wild type, whereas the double mutant displayed a much-pronounced growth defect (Khare et al. 2017). Moreover, all the mutants showed severely compromised growth when iron was added to H_2O_2 containing medium (Khare et al. 2017). These observations also correlated with previous findings wherein the deletion of the transcription regulator IdeR, which controls the

expression of *bfrA* and *bfrB*, also showed increased sensitivity to both H_2O_2 and plumbagin (Rodriguez et al. 2002; Gold et al. 2001). Moreover, BfrA of *P. aeruginosa* and *Brucella abortus* have been shown to impart protection against the H_2O_2, which is mediated by quenching of the iron molecules present in the free form (Ma et al. 1999; Almirón and Ugalde 2010). In *S. typhimurium* and *Campylobacter jejuni*, mutants deficient in the synthesis of Bfr and ferritin, respectively, showed significantly higher sensitivity to H_2O_2 and paraquat when compared with the parental strain (Velayudhan et al. 2007; Wai et al. 1996). These findings in *M. tuberculosis* and other organisms substantiate the importance of Bfrs and ferritins in storing the excess iron and protecting from intracellular iron overload.

17.4.2 Diverse Roles of *M. tuberculosis* BfrA and BfrB

There are some organisms that produce one kind of iron storage proteins, while there are others that express both Bfrs as well as ferritins. *E. coli* has four genes dedicated to iron storage; *bfd* (bacterioferritin-associated ferredoxin), *bfr* (bacterioferritin), *ftnA*, and *ftnB* (Abdul-Tehrani et al. 1999). Majorly, these proteins act as iron reservoirs and release the stored iron in the conditions of iron scarcity, while in iron excess, they quench the excess iron. In *E. coli*, FtnA is the major iron storage protein, while Bfd and Bfr are proposed to play a role in the iron release under low iron conditions (Abdul-Tehrani et al. 1999). However, FtnB is considered to play a diverse role than iron storage as it lacks the ferroxidase catalytic center (Abdul-Tehrani et al. 1999). In another study employing a ferritin-deficient mutant of *Porphyromonas gingivalis*, it was shown that the mutant exhibited slow growth under low iron conditions indicating that ferritins are required to cater to low iron stress by releasing its stored iron, which can be utilized for iron requiring cellular processes (Ratnayake et al. 2000). However, in *P. gingivalis*, the ferritin molecule does not play a role in preventing the cells from oxidative damage since the ferritin mutant grew normally under conditions of oxidative stress (Ratnayake et al. 2000). Similarly, in the case of *Brucella melitensis*, the deletion of the Bfr gene showed no growth defect of the mutant strain in its ability to grow in human macrophages (Denoel et al. 1997). Moreover, among the two iron storage proteins of *E. coli* and *C. jejuni*, it was observed that only ferritin was found to be important for the growth of the bacteria under iron starvation, whereas for the protection of the cells against oxidative damage both ferritin and Bfr were required (Abdul-Tehrani et al. 1999; Wai et al. 1996). Gonococcal Bfr is an important source of iron under iron deprivation and is also involved in protection from oxidative stress (Chen and Morse 1999). In *Erwinia chrysanthemi*, the ferritin protein was found to be important under both iron-deficient as well as oxidative stress conditions (Boughammoura et al. 2008). All these studies suggest that there are diverse and distinct roles assigned to different types of ferritins.

Similar to many other organisms, there are two iron storage proteins in *M. tuberculosis*, namely a heme-bound bacterioferritin (BfrA) and ferritin (BfrB). In a study by Khare et al., gene deletion mutants of these two proteins were

employed to understand their physiological role in *M. tuberculosis* (Khare et al. 2017). Immunoblot analysis showed that iron levels do not influence the levels of cellular BfrA, whereas the amount of BfrB increases when exposed to increasing iron concentrations (Khare et al. 2017). These results also corroborate with the previous findings of the presence of two promoters for the regulation of *bfrA* in response to varying iron levels (Gold et al. 2001). Under conditions of iron scarcity, P_{low} promoter is activated, which drives the expression of BfrA, while under high iron conditions, the expression of BfrA is driven by the other promoter, i.e., P_{high} (Gold et al. 2001). The differential expression of these two iron storage proteins implicates the possible differences in their physiological requirement by *M. tuberculosis* in maintaining iron homeostasis (Khare et al. 2017). Besides, the presence of two different promoters to drive the synthesis of BfrA for maintaining its cellular levels irrespective of iron concentrations indicates its importance in iron homeostasis (Khare et al. 2017). Moreover, the increase in the expression of BfrB under iron excess conditions indicates its importance to quench free cellular iron and protecting the cells from iron-mediated oxidative damage. It was observed that the Δ*bfrA* and Δ*bfrB* double mutant was compromised in its ability to grow under iron starvation. In contrast, single mutants did not show any growth defect suggesting that the presence of either of these proteins can compensate for the loss of the other protein (Khare et al. 2017). Interestingly, the study by Khare et al. showed that under iron starvation, the BfrA levels in the Δ*bfrB* mutant and the parental strain were similar, which suggest that the normal levels of BfrA were sufficient to cater for BfrB deficiency, without any adverse effect on the growth of BfrB mutant. Whereas the levels of BfrB were found to be 1.5 times more in *bfrA* mutant than in the wild type indicating that in the absence of BfrA, normal levels of BfrB could not support normal growth thus, induction of BfrB was required to compensate for the absence of BfrA (Khare et al. 2017). These findings revealed a role of BfrA to serve as an iron reservoir under iron-limiting conditions. Khare et al. also showed that BfrA could release the iron three times faster than BfrB (Khare et al. 2017). The release of iron is accelerated by the presence of heme, which is proposed to be involved in the reduction of the stored ferric ions into ferrous form, which is made available to the cell for its iron requirements (Khare et al. 2017). Indeed, BfrA of *M. tuberculosis* was shown to release the iron in a heme-dependent manner. It was shown that the association of Bfd and Bfr in *E. coli* facilitates iron influx and efflux into the core of the Bfr proteins (Rivera 2017). Moreover, in the case of *P. aeruginosa*, Bfd protein and ferredoxin reductase are known to get upregulated ~200 folds and ~3 folds, respectively, under low iron conditions (Weeratunga et al. 2010; Wang et al. 2007). It is proposed that Bfd protein accepts electrons from the ferredoxin reductases to reduce iron mineral core to release ferrous irons under conditions of iron starvation. Additionally, in vitro studies and X-ray crystallography data also showed the interaction of Bfd protein with recombinant BfrB of *P. aeruginosa* (Eshelman et al. 2017). Further, the release of ferrous iron was shown to be mediated by the presence of heme moiety (Yao et al. 2012, 2011; Weeratunga et al. 2009). Moreover, in the case of *M. tuberculosis*, the expression of the *bfd* gene, present divergent to the *bfrA* gene, is regulated by IdeR (Gold et al. 2001). It was also shown that the

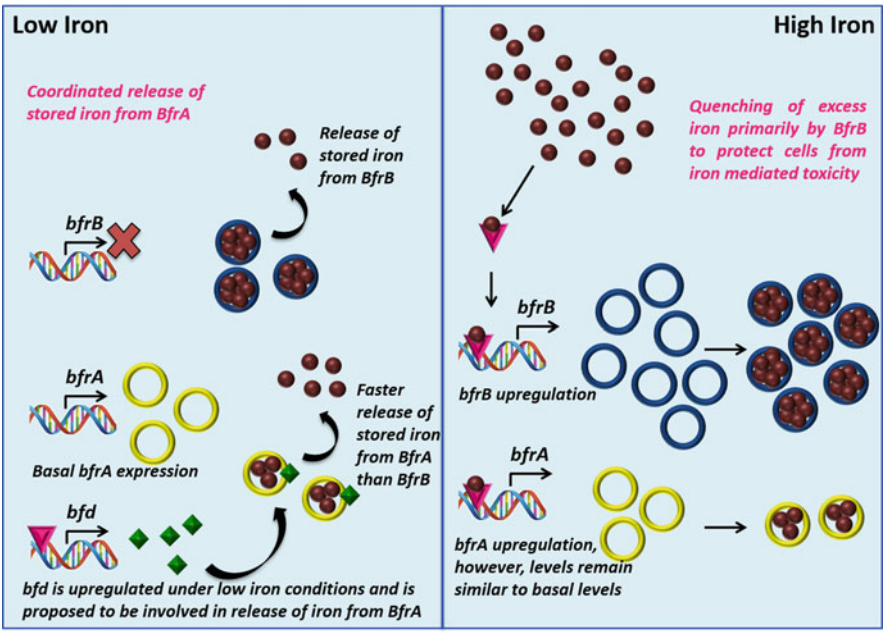


Fig. 17.4 Schematic representation of the model proposed for the differential roles of *M. tuberculosis* iron storage proteins in iron homeostasis. It depicts that under low iron conditions, BfrA plays a vital role in the release of stored iron. In contrast, under iron-rich conditions, BfrB is the preferred protein to quench the excess iron molecules and prevent the cell from oxidative damage. The release of iron is proposed to be accelerated by the presence of heme moiety and/or Bfd proteins

expression level of *bfd* is upregulated under iron limitation; however, the involvement of Bfd in the release of iron has not been shown so far (Gold et al. 2001). Khare et al. also showed that BfrB could store the iron with much higher capacity as compared to BfrA, whereas the latter was more equipped to release its stored iron at a faster rate (Khare et al. 2017). These findings gave important insights into the diverse functions of these proteins, with BfrA playing an important role in iron scarcity, while BfrB acts as a major protein to cater to high iron stress (Fig. 17.4).

Moreover, in a recent study published by Mohanty et al., the mechanism of the ferroxidase activity of *M. tuberculosis* BfrA has been investigated. Their data revealed that the initial rate for the ferroxidase activity of *M. tuberculosis* BfrA exhibits a sigmoidal behavior. BfrA utilizes H_2O_2 in the presence of O_2 to generate the transient intermediate for the formation of ferrihydrite biomineral. In doing so, BfrA also detoxifies H_2O_2, which is the primary agent produced by the host to induce oxidative stress (Mohanty et al. 2019). Since *M. tuberculosis* lacks the genes for Dps and Dps-like proteins, which are involved in protecting the DNA from oxidative damage, the group also investigated whether BfrA exhibits any catalase-like activity, similar to Dps protein (Mohanty et al. 2019). Their results showed that BfrA indeed was able to protect plasmid DNA from oxidative stress agents like Fe^{2+}

and H_2O_2 by forming the complex with plasmid DNA (Mohanty et al. 2019). Thus, BfrA performs multiple functions, including iron release under iron-limiting conditions, detoxification, and catalase-like activity, which are relevant for the pathogenesis of *M. tuberculosis* (Mohanty et al. 2019; Khare et al. 2017). Also, a study by Parida et al. showed that *M. tuberculosis* BfrB also forms transient intermediate species similar to BfrA; however, BfrB could relatively form and decay the transient species much rapidly and faster in comparison to BfrA (Parida et al. 2020). The differences in the amino acids lining the ferroxidase center of both proteins could explain the rapid kinetics of iron entry and oxidation by BfrB (Parida et al. 2020). This also corroborates with earlier finding that *M. tuberculosis* BfrB is the major protein responsible for protecting the cell from iron-mediated oxidative damage by rapidly quenching the free iron, thereby preventing Fenton reaction and subsequent formation of oxygen radicals (Khare et al. 2017).

17.4.3 Ferritins and Hypoxia

There is a unique connection observed between ferritins and hypoxia. In the case of cancer cells and oligodendrocytes, ferritins have been shown to get upregulated in response to hypoxia (Qi et al. 1995; Smith et al. 2003). In patients with many different respiratory disorders, including inflammatory conditions, an increased ferritin concentration was found (Knovich et al. 2009). It is hypothesized that hypoxic conditions can reduce the pH, which in turn affects the iron-sulfur clusters leaching their iron and thereby increasing the intracellular toxic iron. To this end, BfrB of *M. tuberculosis* has been known to be upregulated in response to hypoxia (Rosenkrands et al. 2002; Sherman et al. 2001). *M. tuberculosis* is exposed to hypoxic as well as acidic stress conditions in the host and requires ferritins for circumventing the stressful conditions (Rustad et al. 2009). Khare et al. attempted to understand the role of BfrB in hypoxic conditions and the survival of the pathogen, by monitoring the growth of various *M. tuberculosis bfr* mutants under hypoxic conditions (Khare et al. 2017). It was observed that when subjected to hypoxic conditions, both the single mutants (M.tbΔ*bfrA* and M.tbΔ*bfrB*) exhibited growth defect in comparison to wild-type strain; however, the effect was more pronounced in M.tbΔ*bfrB* mutant. Besides, the mutant lacking both the *M. tuberculosis* ferritins (M.tbΔ*bfrAbfrB*) did not survive at all (Khare et al. 2017). Their study concluded that *M. tuberculosis* iron storage proteins play an essential role in evading the hypoxic conditions.

17.4.4 In Vivo Significance of Mycobacterial BfrA and BfrB

Iron storage proteins are involved in the virulence of many pathogens (Ma et al. 1999; Abdul-Tehrani et al. 1999; Reddy et al. 2012; Chen and Morse 1999; Boughammoura et al. 2008). In a study by Reddy et al., *M. tuberculosis* strain deficient in both BfrA and BfrB proteins exhibited severe attenuation when

compared with the parental strain in human macrophages, suggesting that these proteins play an essential role in the survival of the pathogen under an iron-restrictive environment of activated macrophages (Reddy et al. 2012). The guinea pig studies substantiated the importance of these iron storage proteins in the pathogenesis of *M. tuberculosis*. The double mutant M.tbΔ*bfrAbfrB* exhibited a significantly reduced bacillary load in the spleen of guinea pigs infected with the double mutant in comparison to the bacillary load observed in the spleen of guinea pigs infected with the parental strain when observed at 10 weeks post-infection (Reddy et al. 2012). Moreover, on analyzing the bacillary load at a later time point of the disease progression, i.e., at 16 weeks post-infection, M.tbΔ*bfrAbfrB* exhibited a further reduction in the CFU (Reddy et al. 2012). Thus, both BfrA and BfrB were required for the virulence of *M. tuberculosis* in the guinea pig model (Reddy et al. 2012). BfrA and BfrB proteins were also shown to be involved in the hematogenous spread of the pathogen (Reddy et al. 2012). Since these proteins contribute to bacterial virulence, they appear to be attractive drug targets.

17.5 Insights into the Structural Features of the Ferritin Family of Proteins

There are several crystal structures available in the literature for the proteins belonging to the ferritin family of proteins. The various structures include ferritins from the horse, frog, human, insect, soybean, *E.coli*, *Archaeglobus fulgidus*, *Helicobacter pylori, frog, Pyrococcus furiosus, Listeria innocua, Desulfovibrio desulfuricans, Brevibacillus brevis, horse*, and many others (Lawson et al. 1991; Toussaint et al. 2007; Ha et al. 1999; Trikha et al. 1995, 1994; Granier et al. 1997; Michaux et al. 1996; Masuda et al. 2010; Hamburger et al. 2005; Stillman et al. 2001; Tatur et al. 2007; Johnson et al. 2005; Cho et al. 2009; Ilari et al. 2000; Macedo et al. 2003; Ren et al. 2003). Though there are differences in the sequence of ferritin family of proteins from various species, their overall tertiary and quaternary structures exhibit striking similarities. A single subunit of this class of proteins has a typical 4-helical bundle comprising of helices A–D with a long loop linking helices B and C (Ebrahimi et al. 2015; Honarmand Ebrahimi et al. 2015; Arosio et al. 2017). However, for the function of iron storage, many such subunits assemble to form a spherical shell with an iron storage cavity inside (Ebrahimi et al. 2015; Honarmand Ebrahimi et al. 2015; Arosio et al. 2017). In the case of ferritins and Bfrs, the functional molecule is made up of 24 subunits arranged in an octahedral 432-symmetry giving rise to a globular protein, which can hold 4000–5000 iron atoms in its cavity inside the sphere (Fig. 17.1). An exception to this arrangement is the ferritin from *A. fulgidus* that is arranged in a tetrahedral 23-symmetry (Johnson et al. 2005). The third subfamily comprising of Dps proteins is different in their arrangement as compared to ferritins and Bfrs, wherein 12 subunits arrange themselves into a tetrahedral 23-symmetry resulting in a low iron storage capacity than the other two subfamilies of proteins. These spherical shells are porous to various small molecules such as iron, oxidants, chelators, and reductants utilizing several

pores and channels spanning the entire surface of the macromolecules, which enable the entry and exit of these molecules (Tosha et al. 2010). Although ferritins and Bfrs have similar quaternary structures, they differ because of the presence of heme moiety in the case of Bfrs.

M. tuberculosis possesses both ferritin (BfrB) and a BfrA. BfrA is classified as a Bfr due to the presence of heme molecules, whereas BfrB lacks the heme molecules and hence is classified as ferritin. The three-dimensional crystal structures of both these proteins are available in the PDB and are elaborated in the following.

17.5.1 Structural Features of BfrB: An Example of a Conserved as Well as a Varied Ferritin

Mycobacterial ferritin BfrB exhibits the characteristic macromolecular assembly, as is displayed by other ferritins. 24 subunits arrange themselves into a cage-like structure with a 432 symmetry. Despite the variation in the primary sequence of *M. tuberculosis* BfrB with various ferritins that range from 11 to 28% identity, the overall structure has a high degree of similarity. However, interestingly, multiple sequence alignment of BfrB sequence with that of other ferritins reveals that the mycobacterial ferritins have an extended C-terminus of ~15 amino acids specific to the mycobacteriaceae (Fig. 17.5) (Khare et al. 2011). The crystal structure of the *M. tuberculosis* BfrB (PDB code: 3QD8) was solved by molecular replacement employing the macromolecular 24 meric structure of archaeal *Thermotoga maritima* ferritin (PDB ID: 1vlg) (Khare et al. 2011). As stated above, the structure exhibited high similarity to the structures of other ferritins and Bfrs, with four helices (A–D) and a small E-helix forming a single subunit of the four-helical bundle and many such subunits arranging themselves giving rise to two-, three-, and fourfold axes (Fig. 17.2) (Khare et al. 2011). The extended C-terminus, present uniquely in the mycobacterial ferritins, is referred to as the C (flexible), due to its existence as a less structured loop. It did not show any resemblance to other structures; however, the structure was modeled in one of the subunits resulting in interesting insights (Fig. 17.5) (Khare et al. 2011). This extended C-terminus is involved in imparting stability to the protein and in the iron oxidation property of the protein (Khare et al. 2011). The functional activity of ferritins is to oxidize the ferrous ions into the ferric at the ferroxidase center and further steer these ferric ions inside the cavity where they are stored as the ferric mineral. This initial step involves firstly the uptake of iron and oxidation at the ferroxidase center, which comprises of two iron-binding sites surrounded by acidic residues. One of the most important features for the proper functioning of ferritins is the presence of various electrostatic gradients (Khare et al. 2011). It has been found that the interior surface of the shell predominantly has a negative potential. In contrast, the exterior surface having various channels and pores formed by the assembly of subunits appears to be a mix of both positive and negative (Khare et al. 2011).

a

```
M.bovis          EEVALMATLVRVADRAGAN---------LFELENFVAREVDVAPAASGAPHAAGGRL
M.kansasii       EEMSLMSTLLRIADRAGSN---------LFELENFVAREVGRAPSASGAPRAAGGSL
M.ulcerans       EEVALMASLVRIADRAGAN---------LFELENFVAREVGRAPSGAAAPRAAGGNL
M.marinum        EEVALMASLVRIADRAGAN---------LFELENFVAREVGRAPSGAAAPRAAGGNL
M.smeg           EEVAAMTTLVRIADRAGSN---------LFHIEDFVAREMSAAGADPTAPRAAGGAL
M.abscessus      EEVALMDTLLTVAERAGDN---------YFDLEEFVAREISVVESDPTAPPAAGGAL
H37Ra            EEVALMATLVRVADRAGAN---------LFELENFVAREVDVAPAASGAPHAAGGRL
M.tb_BfrB        EEVALMATLVRVADRAGAN---------LFELENFVAREVDVAPAASGAPHAAGGRL
2CEI             KAIKELGDHVTNLRKMGAP--------ESGLAEYLFDKHTLGDSDNES---------
1R03             KSIKELGDHVHNLVKMGAP--------DAGLAEYLFDTHTLGNENKQN---------
1RCC             ETIKKLGDHITSLKKLWSS--------HPGMAEYLFNKHTLG----------------
1MFR             KDIKRIGDFITNLKRLGLP--------ENGMGEYLFDKHSVKESS-------------
1IES             KLIKKMGDHLTNIQRLVG--------SQAGLGEYLFERLTLKHD--------------
1H96             KLIKKMGNHLTNLRRVAGPQPAQTGAPQGSLGEYLFERLTLKHD--------------
3A68             EAIKRISEYVAQLRRVGKG-----------HGVWHFDQMLLHEGGDAA---------
3BVE             EEEVLFKDILDKIELIGNE----------NHGLYLADQYVKGIAKSRKS---------
1KRQ             EEEALFRGIVDKIKLIGEH----------GNGLYLADQYIKNIALSRKK---------
1EUM             EEEKLFKSIIDKLSLAGKS----------GEGLYFIDKELSTLDTQN-----------
1VLG             EEEDQVREILDLLEKANGQ----------MSVIFQLDRYLGQRE--------------
1S3Q             EEEASALDIVEKLRLIGED----------KRALLFLDKELSLRQFTPPAEEEK-----
1Z6O             EKIRTLAGHTSDLKKFITANNG----HDLSLALYVFDEYLQKTV--------------
2Z90             EDPSTADLLHGLIDGLEKQ--------------AWLIRSENRKV---------------
3BKN             -HIDYLETQLQLMDKLGDA----------LYAAQCVSRPPGSA---------------
2WTL             -HIDYLETQLELMDKLGEE----------LYSAQCVSRPPTSAWSHPQFEK------
```

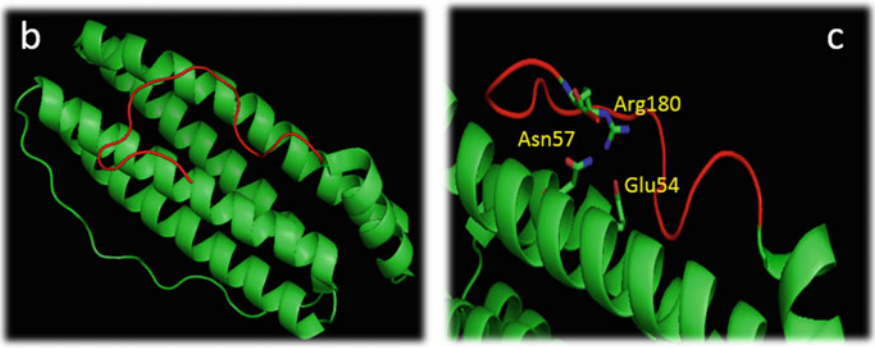


Fig. 17.5 Presence of an extended C-terminus in mycobacterial ferritins. (**a**) Multiple sequence alignment of *M. tuberculosis* BfrB (blue) with various other mycobacterial ferritins (above BfrB) and some other ferritins whose structures are known in PDB (below BfrB). (**b**) The extended C-terminus is shown as a red loop in *M. tuberculosis* BfrB structure, wherein it lies at the interior side of the ferritin molecule facing the cavity. (**c**) The extended C-terminus residue Arg180 forms crucial interactions with Asn57 and Glu54 of the B-helix at the back of the ferroxidase center

17.5.2 Structural Features of BfrA

M. tuberculosis BfrA belongs to the class of Bfrs having a non-covalently bound heme moiety. Two crystal structures have been solved for the *M. tuberculosis* BfrA; the first structure was a selenomethionyl analog of BfrA with a demetalled and degraded heme molecule since the electron density for intact heme was not observed (PDB ID: 2WTL). The second structure was of a heme reconstituted BfrA, wherein heme was exogenously reconstituted to look into the spatial arrangement of heme

binding in BfrA (PDB ID: 3UOI) (Gupta et al. 2009; McMath et al. 2013). As is the case with most other ferritins and Bfrs, *M. tuberculosis* BfrA also exhibits the formation of a spherical macromolecular structure formed by four-, three-, and twofold symmetrical arrangement of 24 individual subunits. Each subunit comprises of four helices followed by a short E-helix at the C-terminus (Gupta et al. 2009). The heme-binding pocket lies between two subunits of a dimer, and sulfur atoms of methionine residues from each of these subunits are involved in the hexacoordination with the heme moiety resulting in a total of 12 heme molecules bound to the complete Bfr molecule (McMath et al. 2013). The conserved di-iron center binds two iron ions similar to other ferritins/Bfrs. The structures depict the presence of several channels at the exterior surface that provides space to various molecules, including iron, protons, and reductants to traverse inside the inner cavity (Gupta et al. 2009).

17.6 Critical Residues of *M. tuberculosis* BfrB

Important residues in various other ferritin proteins have been identified, which are involved in iron storage, iron entry, ferroxidase center activity, and iron mineralization (Lawson et al. 1989; Stillman et al. 2003; Wade et al. 1991; Treffry et al. 1993; Theil et al. 2000; Takagi et al. 1998; Jin et al. 2001; Takahashi and Kuyucak 2003; Kilic et al. 2003; Zhang et al. 2010). For example, in general, many studies through structural and functional analyses have shown the participation of glutamate residues lining the ferroxidase center to participate in iron oxidation, the major step of iron storage (Lawson et al. 1989; Stillman et al. 2003). For instance, in the case of human H-chain ferritin, E62 and H65 of ferroxidase center were shown to be involved in iron-binding. Many studies have shown that the residues lining the threefold axis of the ferritin protein are involved in the entry of the iron (Wade et al. 1991). An increase in the iron release rate was observed when leucine residue lying at the threefold channel of recombinant frog H-ferritin was substituted to proline at the 134th position (Takagi et al. 1998; Jin et al. 2001). The mutation of leucine to proline at the 134th position might result in localized unfolding at the site, which could hamper some important interactions required for the formation of the threefold channel, thereby regulating the iron release (Takagi et al. 1998; Jin et al. 2001). Moreover, few studies have shown the importance of residues at the fourfold channel to be involved in iron exit (Takahashi and Kuyucak 2003). Besides, very few studies have identified residues crucial for protein stability and assembly (Kilic et al. 2003; Zhang et al. 2010). The double mutant of *Rhodobacter capsulatum* Bfr was constructed by site-directed mutagenesis of the two glutamate residues at the threefold axis, which led to the disruption of its 24-mer structure (Kilic et al. 2003). Also, stable dimers rather than 24-mer assembled protein were formed on mutating residues at the threefold channel, suggesting their involvement in assembly formation of *E. coli* Bfr (Zhang et al. 2010). Similarly, in a study by Khare et al., various residues of *M. tuberculosis* BfrB were mutated, and their effect on the structure, assembly, and functionality of the protein was studied (Khare et al. 2013). When

residues of *M. tuberculosis* BfrB at the three- and fourfold axes were mutated, the protein failed to form the characteristic 24-mer assembly and existed as subunit dimers (Khare et al. 2013). R69 forms a salt bridge with D118 as observed by analyzing the BfrB structure, which suggested the role of this intra-subunit interaction in assembly formation, which was substantiated by the fact that R69A mutant existed as stable dimers instead of a 24-mer structure. The importance of the cognate Arg residue has also been shown in other homologous ferritins. L129 and L120 of M. *tuberculosis* BfrB are involved in the formation of an intersubunit hydrophobic contact, which plays an important role in stabilizing the 24-mer structure as L129P mutant failed to assemble into the 24-mer (Khare et al. 2011). The study by Khare et al. also revealed the role of residues present at the fourfold channel of *M. tuberculosis* BfrB in assembly formation. π interactions between F159 and F154 from two adjacent subunits are essential in the oligomerization of the protein in addition to a salt bridge between residues E157 and R162, and these interactions are crucial for the structural integrity of the *M. tuberculosis* BfrB (Khare et al. 2013). *M. tuberculosis* BfrB structure contains two metal-binding sites at the ferroxidase center, metal A and metal B binding site (Khare et al. 2013). It was found that mutating the residues E22, Y29, and H58 present at the ferroxidase center led to a reduction in the iron oxidation rate with the most pronounced effect in case of E22 mutation (Khare et al. 2013). Although few ferritins have two metal-binding sites (A and B), there are a few examples that possess an additional third site C (Toussaint et al. 2007; Stillman et al. 2001, 2003; Treffry et al. 1998). The structure of *M. tuberculosis* BfrB revealed the absence of a third metal-binding site, due to the flipping of E54 (generally present at the third metal-binding site) in the opposite orientation of iron-binding (Khare et al. 2013). Interestingly, mutation of another residue E135 (cognate residue of the third metal-binding site in other homologs) resulted in an increased rate of iron oxidation when compared to the wild-type protein (Khare et al. 2013). Also, the iron incorporation capacity of this mutant was also decreased. Moreover, the mutant protein exhibited iron-induced aggregation along with an enhanced iron oxidation rate. These findings implicated the role of E135 in the iron translocation and migration to the core (Khare et al. 2013).

17.7 Ferritins as Nature's Nanocages

Ferritins are majestic tools for nanofabrication and represent splendid examples of complex architecture designed by nature, which self-assemble into hollow symmetrical protein cages and act as biological containers. The natural biological function of nanocage ferritins to mineralize iron inside their cavity has led to their use in diverse nanotechnological applications as a constrained vessel for various metals and drugs. Several studies have been carried out to understand the role of different amino acid residues of ferritins, which provide a greater understanding about the assembly and function of these proteins, and substitutions of various single amino acid residues can serve as a remarkable tool to provide improved properties to these ferritins for their development into superior nanocages with desirable properties for applications

in drug delivery, bioimaging, MRI, and fabrication of metal oxide semiconductors, etc. Improved knowledge of the factors mediating these self-assembling molecular properties holds great promise for the development of novel tools for nanotechnology (Zhang 2003; Lin et al. 2005). Hikono et al. employed ferritins for the fabrication of quantum nanodot arrays by using their biomineralization property in a process named "bio-nano process," which can be useful for developing nanoscale electronic devices such as transistors, solar cells, and LEDs (Hikono et al. 2003). With the use of doxorubicin and daunomycin encapsulated ferritins, these nanocages are useful for drug delivery. The use of radioactive metal ions-loaded human ferritins, possessing positron emission tomography functionality, in high sensitivity tumor imaging has already been demonstrated (Simsek and Kilic 2005; Ma-Ham et al. 2011; Lin et al. 2011). The drug delivery systems have seen a great deal of advancement since the age of liposome-based carriers to protein-based nanoparticles. The classical liposome-based delivery systems suffered from several caveats such as poor drug encapsulation, thermodynamic instability leading to early drug release, and limited reproducibility (Maham et al. 2009). Most of these disadvantages were addressed by the advent of protein and non-protein-based nanomedicine platforms (Maham et al. 2009). Furthermore, the protein-based delivery systems were superior to the non-protein-based systems due to their bioavailability, biocompatibility, and biodegradability, in addition to reduced cytotoxicity (Maham et al. 2009). The pH-based disassembly and assembly of ferritins have been exploited to encapsulate the cargos inside the ferritin shells. At low pH (pH 2), the octahedral ferritin molecules disassemble into subunits. The cargo is entrapped inside when the ferritin molecules again assemble at basic pH (pH 8.5) (Simsek and Kilic 2005), and Xing et al. has successfully demonstrated this by encapsulating platinum-based anticancer drugs in the horse spleen ferritin cages (Xing et al. 2009). In addition to being a lucrative example for drug delivery systems, ferritins have also shown promise in several other nanotechnological applications. The ability of ferritin cages to mineralize metals has been exploited to enclose paramagnetic gadolinium (Gd^{3+}) having high relativity values, which acts as a contrast agent in MRI (Makino et al. 2011). The use of high temperatures can often accelerate mineralization reactions and lead to higher metal load, which would further enhance the potency of the imaging, for which, generation of thermostable ferritins is desirable. There are numerous electronic applications of ferritins as well, which are primarily based again on the mineralization ability of the ferritins and the uniform nanometer size of these cages. The metal-oxide-loaded ferritins form symmetrical arrays on silicon wafers, which are then treated with ozone to remove the ferritin shells resulting in high-density metal cores in the nanoscale dimensions (Hikono et al. 2003). Such quantum nanodot arrays are necessary to make nanoscale electronic devices such as transistors, solar cells, and LEDs (Hikono et al. 2003).

Thus, ferritins and their variants with more robust and improved properties are beneficial in designing superior nanoparticles that are useful in a wide range of applications, from biomedicine to electronics.

17.8 Conclusions

Iron is an essential metal for all living organisms to survive and carry out important metabolic cellular processes. As much as it is required for various essential physiological functions, its versatile nature demands maintenance of its levels since high levels of iron can undergo Fenton reaction leading to the generation of toxic oxygen intermediates, which can have a deleterious effect on the cells. The iron limitation also poses a significant challenge for pathogenic organisms such as *Mycobacterium tuberculosis*. Pathogens have devised iron acquisition and iron storage proteins to maintain their cellular iron levels. Iron storage proteins had a long history when they were first crystallized in 1937. These proteins have a characteristic 24-meric oligomeric assembly with an inside cavity for the storage of thousands of iron molecules in a mineralized form. These proteins play an important physiological role by acting as a source of iron under conditions of iron scarcity and as efficient iron quenchers under iron excess. *M. tuberculosis* possesses two kinds of iron storage proteins, namely BfrA (heme bound) and non-heme-bound ferritin BfrB. Numerous studies have shown the involvement of these iron storage proteins in the survival of the pathogen under varying iron conditions as well as stressful conditions like oxidative and hypoxic stress. Importantly, both these proteins have been demonstrated to be necessary for the survival of *M. tuberculosis* inside macrophages as well as in guinea pigs. Moreover, these proteins have been shown to perform differential functions under varying iron conditions, with BfrA playing an important role in iron scarcity and BfrB acting as the major protein to cater to high iron stress. Structure of *M. tuberculosis* BfrB showed an extended C-terminus, which was uniquely present in the mycobacterial ferritins, which was found to be involved in imparting stability to the protein and also contributing in its iron oxidation property. Various amino acid residues of *M. tuberculosis* BfrB have been identified by site-directed mutagenesis to be important for oligomerization, macromolecular assembly, and ferroxidase activity. A few mutants identified with improved properties can serve as useful tools for the development of superior ferritin nanocages that could be employed in various applications such as drug discovery, bioimaging, MRI, and fabrication of high-quality metal oxide semiconductors.

References

Abdul-Tehrani H, Hudson AJ, Chang YS, Timms AR, Hawkins C, Williams JM, Harrison PM, Guest JR, Andrews SC (1999) Ferritin mutants of Escherichia coli are iron deficient and growth impaired, and fur mutants are iron deficient. J Bacteriol 181(5):1415–1428

Almirón MA, Ugalde RA (2010) Iron homeostasis in Brucella abortus: the role of bacterioferritin. J Microbiol 48(5):668–673

Arosio P, Adelman TG, Drysdale JW (1978) On ferritin heterogeneity. Further evidence for heteropolymers. J Biol Chem 253(12):4451–4458

Arosio P, Elia L, Poli M (2017) Ferritin, cellular iron storage and regulation. IUBMB Life 69 (6):414–422

Boelaert JR, Vandecasteele SJ, Appelberg R, Gordeuk VR (2007) The effect of the host's iron status on tuberculosis. J Infect Dis 195(12):1745–1753

Bou-Abdallah F (2010) The iron redox and hydrolysis chemistry of the ferritins. Biochim Biophys Acta 1800(8):719–731

Boughammoura A, Matzanke BF, Böttger L, Reverchon S, Lesuisse E, Expert D, Franza T (2008) Differential role of ferritins in iron metabolism and virulence of the plant-pathogenic bacterium Erwinia chrysanthemi 3937. J Bacteriol 190(5):1518–1530

Boyd J, Oza MN, Murphy JR (1990) Molecular cloning and DNA sequence analysis of a diphtheria tox iron-dependent regulatory element (dtxR) from Corynebacterium diphtheria. Proc Natl Acad Sci USA 87(15):5968–5972

Bullen JJ (1981) The significance of iron in infection. Rev Infect Dis 3(6):1127–1138

Bullen JJ, Ward CG, Wallis SN (1974) Virulence and the role of iron in Pseudomonas aeruginosa infection. Infect Immun 10(3):443–450

Calhoun LN, Kwon YM (2011) Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. J Appl Microbiol 110(2):375–386

Carrondo MA (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. EMBO J 22(9):1959–1968

Cellier MF, Courville P, Campion C (2007) Nramp1 phagocyte intracellular metal withdrawal defense. Microbes Infect 9(14–15):1662–1670

Chen CY, Morse SA (1999) Neisseria gonorrhoeae bacterioferritin: structural heterogeneity, involvement in iron storage and protection against oxidative stress. Microbiology 145 (10):2967–2975

Chiancone E, Ceci P (2010) The multifaceted capacity of Dps proteins to combat bacterial stress conditions: detoxification of iron and hydrogen peroxide and DNA binding. Biochim Biophys Acta 1800(8):798–805

Cho KJ, Shin HJ, Lee JH, Kim KJ, Park SS, Lee Y, Lee C, Park SS, Kim KH (2009) The crystal structure of ferritin from Helicobacter pylori reveals unusual conformational changes for iron uptake. J Mol Biol 390(1):83–98

Chou CJ, Wisedchaisri G, Monfeli RR, Oram DM, Holmes RK, Hol WG, Beeson C (2004) Functional studies of the Mycobacterium tuberculosis iron-dependent regulator. J Biol Chem 279(51):53554–53561

Cianciotto NP (2015) An update on iron acquisition by Legionella pneumophila: new pathways for siderophore uptake and ferric iron reduction. Future Microbiol 10(5):841–851

Cornelis P, Matthijs S, Van Oeffelen L (2009) Iron uptake regulation in Pseudomonas aeruginosa. Biometals 22(1):15–22

De Voss JJ, Rutter K, Schroeder BG, Barry CE (1999) Iron acquisition and metabolism by mycobacteria. J Bacteriol 181(15):4443–4451

De Voss JJ, Rutter K, Schroeder BG, Su H, Zhu Y, Barry CE (2000) The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages. Proc Natl Acad Sci U S A 97(3):1252–1257

Deng Z, Wang Q, Liu Z, Zhang M, Machado AC, Chiu TP, Feng C, Zhang Q, Yu L, Qi L, Zheng J (2015) Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator. Nat Commun 6:7642

Denoel PA, Crawford RM, Zygmunt MS, Tibor A, Weynants VE, Godfroid F, Hoover DL, Letesson JJ (1997) Survival of a bacterioferritin deletion mutant of Brucella melitensis 16M in human monocyte-derived macrophages. Infect Immun 65(10):4337–4340

Doukhan L, Predich M, Nair G, Dussurget O, Mandic-Mulec I, Cole ST, Smith DR, Smith I (1995) Genomic organization of the mycobacterial sigma gene cluster. Gene 165(1):67–70

Dussurget O, Rodriguez M, Smith I (1996) An ideR mutant of Mycobacterium smegmatis has derepressed siderophore production and an altered oxidative-stress response. Mol Microbiol 22 (3):535–544

Ebrahimi KH, Hagedoorn PL, Hagen WR (2015) Self-assembly is prerequisite for catalysis of Fe (II) oxidation by catalytically active subunits of ferritin. J Biol Chem 290(44):26801–26810
Eshelman K, Yao H, Hewage AN, Deay JJ, Chandler JR, Rivera M (2017) Inhibiting the BfrB: Bfd interaction in Pseudomonas aeruginosa causes irreversible iron accumulation in bacterioferritin and iron deficiency in the bacterial cytosol. Metallomics 9(6):646–659
Fetherston JD, Kirillina O, Bobrov AG, Paulley JT, Perry RD (2010) The yersiniabactin transport system is critical for the pathogenesis of bubonic and pneumonic plague. Infect Immun 78 (5):2045–2052
Friedman DB, Stauff DL, Pishchany G, Whitwell CW, Torres VJ, Skaar EP (2006) Staphylococcus aureus redirects central metabolism to increase iron availability. PLoS Path 2(8):e87
Furman M, Fica A, Saxena M, Di Fabio JL, Cabello FC (1994) Salmonella typhi iron uptake mutants are attenuated in mice. Infect Immun 62(9):4091–4094
Gangaidzo IT, Moyo VM, Mvundura E, Aggrey G, Murphree NL, Khumalo H, Saungweme T, Kasvosve I, Gomo ZA, Rouault T, Boelaert JR (2001) Association of pulmonary tuberculosis with increased dietary iron. J Infect Dis 184(7):936–939
Gobin J, Horwitz MA (1996) Exochelins of Mycobacterium tuberculosis remove iron from human iron-binding proteins and donate iron to mycobactins in the M. tuberculosis cell wall. J Exp Med 183(4):1527–1532
Gold B, Rodriguez GM, Marras SA, Pentecost M, Smith I (2001) The Mycobacterium tuberculosis IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. Mol Microbiol 42(3):851–865
Golden CA, Kochan I, Spriggs DR (1974) Role of mycobactin in the growth and virulence of tubercle bacilli. Infect Immun 9(1):34–40
Golonka R, San Yeoh B, Vijay-Kumar M (2019) The iron tug-of-war between bacterial siderophores and innate immunity. J Innate Immun 11(3):249–262
Górska A, Sloderbach A, Marszałł MP (2014) Siderophore–drug complexes: potential medicinal applications of the 'Trojan horse' strategy. Trends Pharmacol Sci 35(9):442–449
Granier T, Gallois B, Dautant A, Langlois d'Estaintot B, Precigoux G (1997) Comparison of the structures of the cubic and tetragonal forms of horse-spleen apoferritin. Acta Crystallogr D Biol Crystallogr 53(5):580–587
Günter-Seeboth K, Schupp T (1995) Cloning and sequence analysis of the Corynebacterium diphtheriae dtxR homologue from Streptomyces lividans and S. pilosus encoding a putative iron repressor protein. Gene 166(1):117–119
Gupta V, Gupta RK, Khare G, Salunke DM, Tyagi AK (2009) Crystal structure of Bfr A from Mycobacterium tuberculosis: incorporation of selenomethionine results in cleavage and demetallation of haem. PLoS One 4(11):e8028
Ha Y, Shi D, Small GW, Theil EC, Allewell NM (1999) Crystal structure of bullfrog M ferritin at 2.8 Å resolution: analysis of subunit interactions and the binuclear metal center. J Biol Inorg Chem 4(3):243–256
Haikarainen T, Papageorgiou AC (2010) Dps-like proteins: structural and functional insights into a versatile protein family. Cell Mol Life Sci 67(3):341–351
Hamburger AE, West AP Jr, Hamburger ZA, Hamburger P, Bjorkman PJ (2005) Crystal structure of a secreted insect ferritin reveals a symmetrical arrangement of heavy and light chains. J Mol Biol 349(3):558–569
Hammer ND, Skaar EP (2011) Molecular mechanisms of Staphylococcus aureus iron acquisition. Annu Rev Microbiol 65:129–147
Hantke K (1981) Regulation of ferric iron transport in Escherichia coli K12: isolation of a constitutive mutant. Mol Gen Genet 182(2):288–292
Hantke K (2001) Iron and metal regulation in bacteria. Curr Opin Microbiol 4(2):172–177
Harrison PM, Arosio P (1996) The ferritins: molecular properties, iron storage function and cellular regulation. Biochim Biophys Acta 1275(3):161–203
Hikono T, Uraoka Y, Fuyuki T, Yamashita I (2003) Novel method for making nanodot arrays using a cage-like protein. Jpn J Appl Phys 42(4A):L398

Honarmand Ebrahimi K, Hagedoorn PL, Hagen WR (2015) Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin. Chem Rev 115(1):295–326

Ilari A, Stefanini S, Chiancone E, Tsernoglou D (2000) The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. Nat Struct Biol 7(1):38

Jin W, Takagi H, Pancorbo B, Theil EC (2001) "Opening" the ferritin pore for iron release by mutation of conserved amino acids at interhelix and loop sites. Biochemistry 40(25):7525–7532

Johnson E, Cascio D, Sawaya MR, Gingery M, Schröder I (2005) Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilic archaeon Archaeoglobus fulgidus. Structure 13 (4):637–648

Keyer K, Imlay JA (1996) Superoxide accelerates DNA damage by elevating free-iron levels. Proc Natl Acad Sci U S A 93(24):13635–13640

Khare G, Gupta V, Nangpal P, Gupta RK, Sauter NK, Tyagi AK (2011) Ferritin structure from Mycobacterium tuberculosis: comparative study with homologues identifies extended C-terminus involved in ferroxidase activity. PLoS One 6(4):e18570

Khare G, Nangpal P, Tyagi AK (2013) Unique residues at the 3-fold and 4-fold axis of mycobacterial ferritin are involved in oligomer switching. Biochemistry 52(10):1694–1704

Khare G, Nangpal P, Tyagi AK (2017) Differential roles of iron storage proteins in maintaining the iron homeostasis in Mycobacterium tuberculosis. PLoS One 12(1):e0169545

Kilic MA, Spiro S, Moore GR (2003) Stability of a 24-meric homopolymer: comparative studies of assembly-defective mutants of Rhodobacter capsulatus bacterioferritin and the native protein. Protein Sci 12(8):1663–1674

Knovich MA, Storey JA, Coffman LG, Torti SV, Torti FM (2009) Ferritin for the clinician. Blood Rev 23(3):95–104

Kochan I (1969) Mechanism of tuberculostasis in mammalian serum. I. Role of transferrin in human serum tuberculostasis. J Infect Dis 1:11–18

Kochan I (1973) The role of iron in bacterial infections, with special consideration of host-tubercle bacillus interaction. In: Arber W et al (eds) Current topics in microbiology and immunology. Springer, Berlin, pp 1–30

Krithika R, Marathe U, Saxena P, Ansari MZ, Mohanty D, Gokhale RS (2006) A genetic locus required for iron acquisition in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103 (7):2069–2074

Laufberger V (1937) Sur la cristallisation de la ferritine. Bull Soc Chim Biol 19:1575–1582

Lawson DM, Treffry A, Artymiuk PJ, Harrison PM, Yewdall SJ, Luzzago A, Cesareni G, Levi S, Arosio P (1989) Identification of the ferroxidase centre in ferritin. FEBS Lett 254(1–2):207–210

Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JM, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesareni G, Thomas CD (1991) Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. Nature 349(6309):541

Levi S, Luzzago A, Cesareni G, Cozzi A, Franceschinelli F, Albertini A, Arosio P (1988) Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. A study of iron uptake and ferro-oxidase activity of human liver, recombinant H-chain ferritins, and of two H-chain deletion mutants. J Biol Chem 263(34):18086–18092

Lin Y, Böker A, He J, Sill K, Xiang H, Abetz C, Li X, Wang J, Emrick T, Long S, Wang Q (2005) Self-directed self-assembly of nanoparticle/copolymer mixtures. Nature 434(7029):55

Lin X, Xie J, Niu G, Zhang F, Gao H, Yang M, Quan Q, Aronova MA, Zhang G, Lee S, Leapman R (2011) Chimeric ferritin nanocages for multiple function loading and multimodal imaging. Nano Lett 11(2):814–819

Lounis N, Truffot-Pernot C, Grosset J, Gordeuk VR, Boelaert JR (2001) Iron and Mycobacterium tuberculosis infection. J Clin Virol 20(3):123–126

Ma JF, Ochsner UA, Klotz MG, Nanayakkara VK, Howell ML, Johnson Z, Posey JE, Vasil ML, Monaco JJ, Hassett DJ (1999) Bacterioferritin A modulates catalase A (KatA) activity and resistance to hydrogen peroxide in Pseudomonas aeruginosa. J Bacteriol 181(12):3730–3742

Macedo S, Romão CV, Mitchell E, Matias PM, Liu MY, Xavier AV, LeGall J, Teixeira M, Lindley P, Carrondo MA (2003) The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans. Nat Struct Biol 10(4):285–290

MaHam A, Tang Z, Wu H, Wang J, Lin Y (2009) Protein-based nanomedicine platforms for drug delivery. Small 5(15):1706–1721

Ma-Ham A, Wu H, Wang J, Kang X, Zhang Y, Lin Y (2011) Apoferritin-based nanomedicine platform for drug delivery: equilibrium binding study of daunomycin with DNA. J Mater Chem B 21(24):8700–8708

Makino A, Harada H, Okada T, Kimura H, Amano H, Saji H, Hiraoka M, Kimura S (2011) Effective encapsulation of a new cationic gadolinium chelate into apoferritin and its evaluation as an MRI contrast agent. Nanomedicine 7(5):638–646

Masuda T, Goto F, Yoshihara T, Mikami B (2010) Crystal structure of plant ferritin reveals a novel metal binding site that functions as a transit site for metal transfer in ferritin. J Biol Chem 285 (6):4049–4059

McMath LM, Contreras H, Owens CP, Goulding CW (2013) The structural characterization of bacterioferritin, BfrA, from Mycobacterium tuberculosis. J Porphyrins Phthalocyanines 17 (03):229–239

Meyer JM, Neely A, Stintzi A, Georges C, Holder IA (1996) Pyoverdin is essential for virulence of Pseudomonas aeruginosa. Infect Immun 64(2):518–523

Michaux MA, Dautant A, Gallois B, Granier T, d'Estaintot BL, Précigoux G (1996) Structural investigation of the complexation properties between horse spleen apoferritin and metalloporphyrins. Proteins 24(3):314–321

Miethke M, Marahiel MA (2007) Siderophore-based iron acquisition and pathogen control. Microbiol Mol Biol Rev 71(3):413–451

Minandri F, Imperi F, Frangipani E, Bonchi C, Visaggio D, Facchini M, Pasquali P, Bragonzi A, Visca P (2016) Role of iron uptake systems in Pseudomonas aeruginosa virulence and airway infection. Infect Immun 84(8):2324–2335

Mohanty A, Subhadarshanee B, Barman P, Mahapatra C, Aishwarya B, Behera RK (2019) Iron mineralizing Bacterioferritin A from Mycobacterium tuberculosis exhibits unique catalase-Dps-like dual activities. Inorg Chem 58(8):4741–4752

Muller D, Edwards ML, Smith DW (1983) Changes in iron and transferrin levels and body temperature in experimental airborne legionellosis. J Infect Dis 147(2):302–307

Neilands JB (1995) Siderophores: structure and function of microbial iron transport compounds. J Biol Chem 270(45):26723–26726

Ochsner UA, Wilderman PJ, Vasil AI, Vasil ML (2002) GeneChip® expression analysis of the iron starvation response in Pseudomonas aeruginosa: identification of novel pyoverdine biosynthesis genes. Mol Micobiol 45(5):1277–1287

Pandey R, Rodriguez GM (2014) IdeR is required for iron homeostasis and virulence in M ycobacterium tuberculosis. Mol Microbiol 91(1):98–109

Parida A, Mohanty A, Kansara BT, Behera RK (2020) Impact of phosphate on iron mineralization and mobilization in nonheme Bacterioferritin B from Mycobacterium tuberculosis. Inorg Chem 59(1):629–641

Qi Y, Jamindar TM, Dawson G (1995) Hypoxia alters iron homeostasis and induces ferritin synthesis in oligodendrocytes. J Neurochem 64(6):2458–2464

Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT (1998) Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem Biol 5(11):631–645

Ratledge C (1999) Iron metabolism. In: Ratledge C, Dale J (eds) Mycobacteria: molecular biology and virulence. Blackwell Science, London, pp 260–280

Ratnayake DB, Wai SN, Shi Y, Amako K, Nakayama H, Nakayama K (2000) Ferritin from the obligate anaerobe Porphyromonas gingivalis: purification, gene cloning and mutant studies. Microbiology 146(5):1119–1127

Reddy PV, Puri RV, Khera A, Tyagi AK (2012) Iron storage proteins are essential for the survival and pathogenesis of Mycobacterium tuberculosis in THP-1 macrophages and the guinea pig model of infection. J Bacteriol 194(3):567–575

Reddy PV, Puri RV, Chauhan P, Kar R, Rohilla A, Khera A, Tyagi AK (2013) Disruption of mycobactin biosynthesis leads to attenuation of Mycobacterium tuberculosis for growth and virulence. J Infect Dis 208(8):1255–1265

Ren B, Tibbelin G, Kajino T, Asami O, Ladenstein R (2003) The multi-layered structure of Dps with a novel di-nuclear ferroxidase center. J Mol Biol 329(3):467–477

Rivera M (2017) Bacterioferritin: structure, dynamics, and protein–protein interactions at play in iron storage and mobilization. Acc Chem Res 50(2):331–340

Rodriguez GM, Smith I (2003) Mechanisms of iron regulation in mycobacteria: role in physiology and virulence. Mol Microbiol 47(6):1485–1494

Rodriguez GM, Smith I (2006) Identification of an ABC transporter required for iron acquisition and virulence in Mycobacterium tuberculosis. J Bacteriol 188(2):424–430

Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I (2002) IdeR, an essential gene in Mycobacterium tuberculosis: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect Immun 70(7):3371–3381

Rosenkrands I, Slayden RA, Crawford J, Aagaard C, Clifton E III, Andersen P (2002) Hypoxic response of Mycobacterium tuberculosis studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. J Bacteriol 184(13):3485–3491

Rustad TR, Sherrid AM, Minch KJ, Sherman DR (2009) Hypoxia: a window into Mycobacterium tuberculosis latency. Cell Microbiol 11(8):1151–1159

Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48(1):77–84

Sauton B (1912) Sur la nutrition minérale du bacille tuberculeux. CR Hebd Séances Acad Sci 155:860–861

Schmitt MP, Holmes RK (1991) Iron-dependent regulation of diphtheria toxin and siderophore expression by the cloned Corynebacterium diphtheriae repressor gene dtxR in C. diphtheriae C7 strains. Infect Immun 59(6):1899–1904

Schmitt MP, Predich M, Doukhan L, Smith I, Holmes RK (1995) Characterization of an iron-dependent regulatory protein (IdeR) of Mycobacterium tuberculosis as a functional homolog of the diphtheria toxin repressor (DtxR) from Corynebacterium diphtheriae. Infect Immun 63 (11):4284–4289

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK (2003) Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J Exp Med 198 (5):693–704

Searle S, Bright NA, Roach TI, Atkinson PG, Barton CH, Meloen RH, Blackwell JM (1998) Localisation of Nramp1 in macrophages: modulation with activation and infection. J Cell Sci 111(19):2855–2866

Sebbane F, Jarrett C, Gardner D, Long D, Hinnebusch BJ (2010) Role of the Yersinia pestis yersiniabactin iron acquisition system in the incidence of flea-borne plague. PLoS One 5(12): e14379

Sheldon JR, Heinrichs DE (2015) Recent developments in understanding the iron acquisition strategies of gram positive pathogens. FEMS Microbiol Rev 39(4):592–630

Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK (2001) Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding α-crystallin. Proc Natl Acad Sci U S A 98(13):7534–7539

Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng TY, Siddiqi N, Fortune SM, Moody DB, Rubin EJ (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc Natl Acad Sci U S A 106(44):18792–18797

Simsek E, Kilic MA (2005) Magic ferritin: a novel chemotherapeutic encapsulation bullet. J Magn Magn Mater 293(1):509–513

Smith JJ, O'Brien-Ladner AR, Kaiser CR, Wesselius LJ (2003) Effects of hypoxia and nitric oxide on ferritin content of alveolar cells. J Lab Clin Med 141(5):309–317

Sritharan M (2016) Iron homeostasis in Mycobacterium tuberculosis: mechanistic insights into siderophore-mediated iron uptake. J Bacteriol 198(18):2399–2409

Stillman TJ, Hempstead PD, Artymiuk PJ, Andrews SC, Hudson AJ, Treffry A, Guest JR, Harrison PM (2001) The high-resolution X-ray crystallographic structure of the ferritin (EcFtnA) of Escherichia coli; comparison with human H ferritin (HuHF) and the structures of the Fe3+ and Zn2+ derivatives. J Mol Biol 307(2):587–603

Stillman TJ, Connolly PP, Latimer CL, Morland AF, Quail MA, Andrews SC, Treffry A, Guest JR, Artymiuk PJ, Harrison PM (2003) Insights into the effects on metal binding of the systematic substitution of five key glutamate ligands in the ferritin of Escherichia coli. J Biol Chem 278 (28):26275–26286

Takagi H, Shi D, Ha Y, Allewell NM, Theil EC (1998) Localized unfolding at the junction of three ferritin subunits a mechanism for iron release? J Biol Chem 273(30):18685–18688

Takahashi T, Kuyucak S (2003) Functional properties of threefold and fourfold channels in ferritin deduced from electrostatic calculations. Biophys J 84(4):2256–2263

Tatur J, Hagen WR, Matias PM (2007) Crystal structure of the ferritin from the hyperthermophilic archaeal anaerobe Pyrococcus furiosus. J Biol Inorg Chem 12(5):615–630

Theil EC, Takagi H, Small GW, He L, Tipton AR, Danger D (2000) The ferritin iron entry and exit problem. Inorg Chim Acta 297(1–2):242–251

Timm J, Post FA, Bekker LG, Walther GB, Wainwright HC, Manganelli R, Chan WT, Tsenova L, Gold B, Smith I, Kaplan G (2003) Differential expression of iron-, carbon-, and oxygen-responsive mycobacterial genes in the lungs of chronically infected mice and tuberculosis patients. Proc Natl Acad Sci U S A 100(24):14321–14326

Tosha T, Ng HL, Bhattasali O, Alber T, Theil EC (2010) Moving metal ions through ferritin−protein nanocages from three-fold pores to catalytic sites. J Am Chem Soc 132 (41):14562–14569

Touati D (2000) Iron and oxidative stress in bacteria. Arch Biochem Biophys 373(1):1–6

Toussaint L, Bertrand L, Hue L, Crichton RR, Declercq JP (2007) High-resolution X-ray structures of human apoferritin H-chain mutants correlated with their activity and metal-binding sites. J Mol Biol 365(2):440–452

Treffry A, Bauminger ER, Hechel D, Hodson NW, Nowik I, Yewdall SJ, Harrison PM (1993) Defining the roles of the threefold channels in iron uptake, iron oxidation and iron-core formation in ferritin: a study aided by site-directed mutagenesis. Biochem J 296(3):721–728

Treffry A, Zhao Z, Quail MA, Guest JR, Harrison PM (1998) How the presence of three iron binding sites affects the iron storage function of the ferritin (EcFtnA) of Escherichia coli. FEBS Lett 432(3):213–218

Trikha J, Waldo GS, Lewandowski FA, Ha Y, Theil EC, Weber PC, Allewell NM (1994) Crystallization and structural analysis of bullfrog red cell L-subunit ferritins. Proteins 18 (2):107–118

Trikha J, Theil EC, Allewell NM (1995) High resolution crystal structures of amphibian red-cell L ferritin: potential roles for structural plasticity and solvation in function. J Mol Biol 248 (5):949–967

Velayudhan J, Castor M, Richardson A, Main-Hester KL, Fang FC (2007) The role of ferritins in the physiology of Salmonella enterica sv. Typhimurium: a unique role for ferritin B in iron-sulphur cluster repair and virulence. Mol Microbiol 63(5):1495–1507

Wade VJ, Levi S, Arosio P, Treffry A, Harrison PM, Mann S (1991) Influence of site-directed modifications on the formation of iron cores in ferritin. J Mol Biol 221(4):1443–1452

Wagner D, Maser J, Lai B, Cai Z, Barry CE, Zu Bentrup KH, Russell DG, Bermudez LE (2005) Elemental analysis of Mycobacterium avium-, Mycobacterium tuberculosis-, and Mycobacterium smegmatis-containing phagosomes indicates pathogen-induced microenvironments within the host cell's endosomal system. J Immunol 174(3):1491–1500

Wai SN, Nakayama K, Umene K, Moriya T, Amako K (1996) Construction of a ferritin-deficient mutant of Campylobacter jejuni: contribution of ferritin to iron storage and protection against oxidative stress. Mol Microbiol 20(6):1127–1134

Wang A, Zeng Y, Han H, Weeratunga S, Morgan BN, Moënne-Loccoz P, Schönbrunn E, Rivera M (2007) Biochemical and structural characterization of Pseudomonas aeruginosa Bfd and FPR: ferredoxin NADP+ reductase and not ferredoxin is the redox partner of heme oxygenase under iron-starvation conditions. Biochemistry 46(43):12198–12211

Weeratunga SK, Gee CE, Lovell S, Zeng Y, Woodin CL, Rivera M (2009) Binding of Pseudomonas aeruginosa apobacterioferritin-associated ferredoxin to bacterioferritin B promotes heme mediation of electron delivery and mobilization of core mineral iron. Biochemistry 48 (31):7420–7431

Weeratunga SK, Lovell S, Yao H, Battaile KP, Fischer CJ, Gee CE, Rivera M (2010) Structural studies of bacterioferritin B from Pseudomonas aeruginosa suggest a gating mechanism for iron uptake via the ferroxidase center. Biochemistry 49(6):1160–1175

Weinberg ED (1975) Nutritional immunity: host's attempt to withhold iron from microbial invaders. JAMA 231(1):39–41

Weinberg ED (1984) Iron withholding: a defense against infection and neoplasia. Physiol Rev 64 (1):65–102

Weinberg ED (1999) Iron loading and disease surveillance. Emerg Infect Dis 5(3):346

Wells RM, Jones CM, Xi Z, Speer A, Danilchanka O, Doornbos KS, Sun P, Wu F, Tian C, Niederweis M (2013) Discovery of a siderophore export system essential for virulence of Mycobacterium tuberculosis. PLoS Path 9(1):e1003120

WHO (2019) Global TB report. https://www.who.int/tb/publications/global_report/tb19_Exec_Sum_12Nov2019.pdf?ua=1. Accessed 25 Jan 2020

Xing R, Wang X, Zhang C, Zhang Y, Wang Q, Yang Z, Guo Z (2009) Characterization and cellular uptake of platinum anticancer drugs encapsulated in apoferritin. J Biol Inorg Chem 103 (7):1039–1044

Yao H, Jepkorir G, Lovell S, Nama PV, Weeratunga S, Battaile KP, Rivera M (2011) Two distinct ferritin-like molecules in Pseudomonas aeruginosa: the product of the bfrA gene is a bacterial ferritin (FtnA) and not a bacterioferritin (Bfr). Biochemistry 50(23):5236–5248

Yao H, Wang Y, Lovell S, Kumar R, Ruvinsky AM, Battaile KP, Vakser IA, Rivera M (2012) The structure of the BfrB–Bfd complex reveals protein–protein interactions enabling iron release from bacterioferritin. J Am Chem Soc 134(32):13470–13481

Zhang S (2003) Fabrication of novel biomaterials through molecular self-assembly. Nat Biotechnol 21(10):1171

Zhang Y, Raudah S, Teo H, Teo GW, Fan R, Sun X, Orner BP (2010) Alanine-shaving mutagenesis to determine key interfacial residues governing the assembly of a nano-cage maxi-ferritin. J Biol Chem 285(16):12078–12086

Correction to: Structural and Functional Aspects of Muscarinic Receptors in Correlation with Anticholinergic Drugs

Pramod Kumar Singh, Rajendra Nath, Ram Naraian, and Manish Kumar Gupta

Correction to:
Chapter 13 in: D. B. Singh, T. Tripathi (eds.),
***Frontiers in Protein Structure, Function, and Dynamics*,**
https://doi.org/10.1007/978-981-15-5530-5_13

The book was inadvertently published with an incorrect affiliation of the authors Ram Naraian and Manish Kumar Gupta in Chapter 13 as Department of Biochemistry, Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India. The affiliation has been corrected as Department of Biotechnology, Veer Bahadur Singh Purvanchal University, Jaunpur, Uttar Pradesh, India.

The updated online version of this chapter can be found at
https://doi.org/10.1007/978-981-15-5530-5_13

D. B. Singh, T. Tripathi (eds.), *Frontiers in Protein Structure, Function, and Dynamics*, https://doi.org/10.1007/978-981-15-5530-5_18