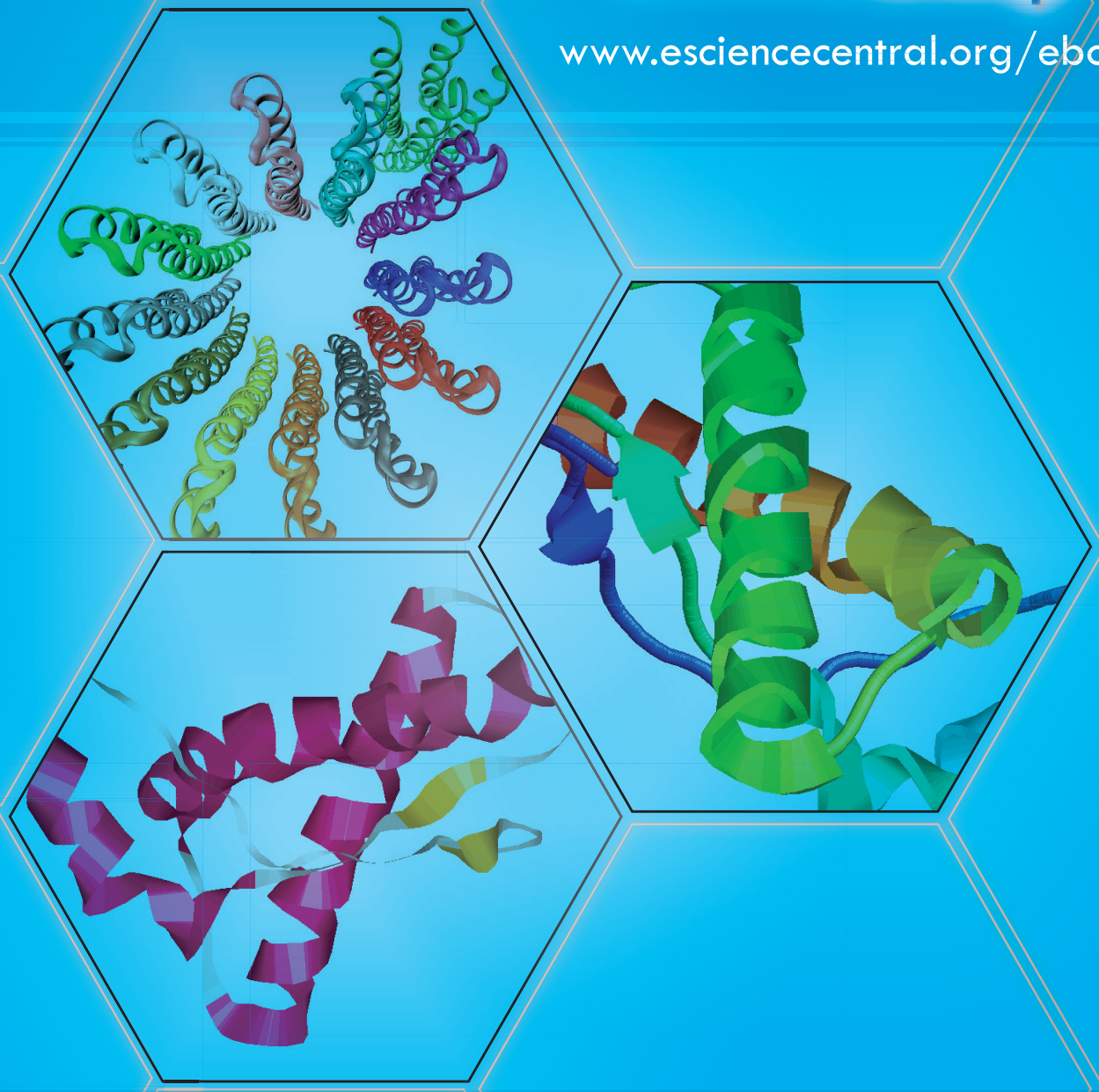
OMICS eBOOKS Group

www.esciencecentral.org/ebooks



Advances in Protein Chemistry

Edited by

Ghulam Md Ashraf

Ishfaq Ahmed Sheikh

Editors

Dr. Ghulam Md Ashraf

(Email: ashraf.gm@gmail.com)

Dr. Ishfaq Ahmed Sheikh

(Email: sheikhishfaq@gmail.com)

King Fahd Medical Research Center

King Abdulaziz University, P.O. Box 80216

Jeddah, Saudi Arabia

Tyrosine Nitrated Proteins: Biochemistry and Pathophysiology

Haseeb Ahsan*

Department of Biochemistry, Faculty of Dentistry, Jamia Millia Islamia (A Central University), Okhla, New Delhi – 110025, India

***Corresponding author:** Dr. Haseeb Ahsan, Department of Biochemistry, Faculty of Dentistry, Jamia Millia Islamia (A Central University), Okhla, New Delhi – 110025, India, E-mail: drhahsan@gmail.com

Abstract

The free radical-mediated damage to proteins results in the modification of amino acid residues, cross-linking of side chains and fragmentation. L-tyrosine and protein bound tyrosine are prone to attack by various mediators and reactive nitrogen intermediates to form 3-nitrotyrosine (3-NT). 3-NT formation is also catalyzed by a class of peroxidases utilizing nitrite and hydrogen peroxide as substrates. Evidence supports the formation of 3-NT *in vivo* in diverse pathologic conditions and 3-NT is thought to be a relatively specific marker of oxidative damage. The formation of nitrotyrosine represents a specific peroxynitrite-mediated protein modification; thus, detection of nitrotyrosine in proteins is considered as a biomarker for endogenous peroxynitrite activity. Formation of tyrosine nitrated proteins is considered to be a post-translational modification with important pathophysiological consequences and is one of the markers of nitrosative stress that have been reported in neurodegeneration, inflammatory and other pathological conditions.

Introduction

Most proteins contain tyrosine residues with a natural abundance of about 3% [1]. Tyrosine (one letter abbreviation, Y; three letter abbreviation, Tyr; also known as 4-hydroxyphenylalanine) is a non-essential amino acid and a member of the aromatic amino acid group. Tyrosine is mildly hydrophilic, a characteristic feature that is explained by the rather hydrophobic aromatic benzene ring carrying a hydroxyl group. As a consequence, tyrosine is often surface-exposed in proteins (only 15% of tyrosine residues are buried inside a protein) and therefore should be available for modification such as nitration by various factors [2-4].

Tyrosine can be modified by the addition of a nitro group (-NO_2) *in vivo* with several agents resulting in the formation of 3-nitrotyrosine (3-NT) or tyrosine nitrated proteins (Figure 1). 3-NT [(2-Amino-3-(4-hydroxy-3-nitrophenyl) propanoic acid)] is a post-translational modification in proteins occurring through the action of a nitrating agent resulting in the addition of a -NO_2 group (in *ortho* position

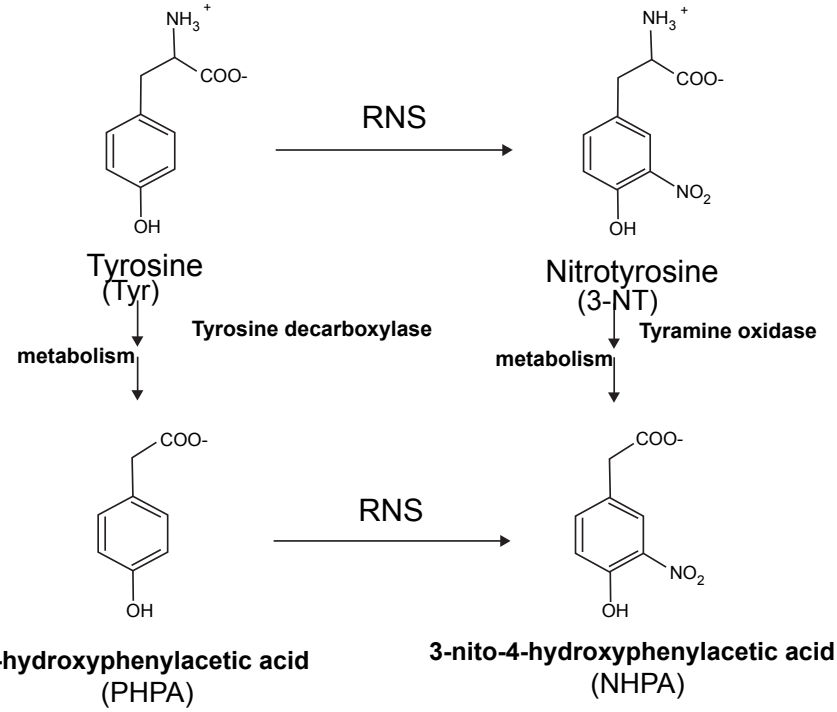


Figure 1: Metabolism of tyrosine leading to the formation of products: p-hydroxyphenylacetic acid and 3-nitro-4-hydroxyphenylacetic acid (*adapted from Mani et al., 2003*).

to the phenolic hydroxyl group) leading to protein tyrosine nitration (PTN) [5]. Protein tyrosine nitration exhibits a certain degree of selectivity and not all tyrosine residues are nitrated since nitration may rather depend on the residue's accessibility to solvents. Neither the abundance of a protein nor the number of tyrosine residues in a given protein can help us predict whether it is a target for PTN [2,3,6]. For example, human serum albumin (HSA) is less extensively nitrated than other plasma proteins, although being the most abundant plasma protein [6]. While HSA has 18 tyrosine residues, an *in vitro* study of peroxynitrite-mediated PTN showed that only two tyrosine residues are particularly susceptible to nitration [7].

The reactivity of a tyrosine residue might also depend on the nature of the reactive species [2]. While peroxynitrite ($ONOO^-$) and tetranitromethane (TNM) nitrate certain proteins [8], there are differences in PTN patterns in other proteins [9-11]. The nitration of protein tyrosine residues could dramatically change protein structure and conformation and subsequently alter their function [12-15]. Tyrosine nitration sites are localized within specific functional domains of nitrated proteins [16]. For example, a strong inhibition of the catalytic activity of manganese-superoxide dismutase (MnSOD) by peroxynitrite-mediated PTN has been reported and explained by nitration of the essential tyrosine residue [17]. The inactivation of human MnSOD by peroxynitrite is caused by exclusive nitration of tyrosine 34 (Tyr34) to 3-nitrotyrosine [18].

Role of Free Radical Species in Tyrosine Nitration

Nitrogen dioxide, nitrous acid, nitryl chloride, and certain peroxidases [19] derived from inflammatory cells can mediate the nitration of tyrosine to form 3-NT (Table 1). For example, nitrite (NO_2-), a primary autoxidation product of NO [20], is further oxidized to form nitrogen dioxide by the action of peroxidases, e.g., myeloperoxidase and eosinophil peroxidase, heme proteins that are abundantly expressed in activated leukocytes. The resulting nitrogen dioxide nitrates the tyrosine residues in the presence of hydrogen peroxide (H_2O_2) [21]. Therefore, tyrosine nitration is based on the generation of nitrogen dioxide radicals (NO_2) by various hemoperoxydases in the presence of H_2O_2 and nitrite [22-24]. Other plausible reactions are based on (i) the interaction of nitric oxide with a tyrosyl radical, (ii) the direct action of nitrogen dioxide, (iii) the formation of nitrous acid by acidification of nitrite, (iv) the oxidation of nitrite by hypochlorous acid to form nitryl chloride (NO_2Cl), (v) the action of acyl or alkyl nitrates or (vi) the action of metal nitrates [1,25,26]. Hence, nitrotyrosine is therefore likely not a footprint for peroxynitrite alone but more generally a marker of nitrative stress.

Nitrating Agent	Proteins nitrated	References
Hemoperoxidases (myeloperoxidase, MPO peroxidase) – MPO/H_2O_2 system	lipoproteins	Tsikas (2012), Abello et al (2009).
Nitrite (NO_2^-)	hemoglobin	
Nitroprusside (*nitropress*)	p65 (NF-κB)	
Peroxynitrite (ONO_2^-)	human serum albumin, creatine kinase, cytochrome c, hemoglobin, NADH dehydrogenase, succinate dehydrogenase, H2A histone protein	
Tetranitromethane (TNM)	bovine serum albumin, human serum albumin	

Table 1: Factors or agents that lead to the formation of tyrosine nitrated proteins.

The reactive nitrogen species, peroxynitrite, is an important nitrating agent *in vivo*. However, it is not the only source of 3-NT formation *in vivo*. Initially, 3-NT was thought to be a biomarker of the existence of peroxynitrite, and protein tyrosine nitration was believed to be the biomarker of peroxynitrite formation in biological systems [27,28]. However, later studies demonstrated that some hemoproteins such as hemoglobin and myoglobin could catalyze $NaNO_2/H_2O_2$-dependent nitration of tyrosine to yield 3-NT [29-31]. The possible underlying mechanism is that hemoproteins catalyze the oxidation of nitrite to nitrogen dioxide (NO_2) which reacts with tyrosine radical (Tyr) to form 3-NT [27,32,33].

Peroxynitrite and related reactive nitrogen species are capable of both oxidation and nitration of the aromatic side-chains of tyrosine and tryptophan in proteins [34,35], resulting in a condition known as "nitrosative stress". Reactive oxygen species (ROS) and reactive nitrogen species (RNS)-mediated damage in the central nervous tissues may reflect an underlying neuroinflammatory process [36]. Protein damage that occurs under conditions of oxidative stress may represent direct oxidation of protein side-chains by ROS and/or RNS or adduction of secondary products of oxidation of sugars (glycoxidation), or polyunsaturated fatty acids (lipid peroxidation) [36]. In addition to the well known or standard reactive oxygen and nitrogen species, oxidative damage to proteins can occur due to alternate oxidants (e.g., HOCl) and circulating oxidized amino acids such as tyrosine radical generated by metalloenzymes such as, myeloperoxidase [37]. The accumulation of oxidized protein is a complex function of the rates of ROS formation, antioxidant levels, and the ability to proteolytically eliminate ion of oxidized forms of proteins [38].

Carbon dioxide/bicarbonate (1.3 and 25 mM in plasma, respectively) [39] strongly influence peroxynitrite-mediated reactions [39-43] and enhance nitration of aromatic rings as in tyrosine. They can also promote nitration in the presence of antioxidants (such as uric acid, ascorbate and thiols), which normally inhibit nitration, while partially inhibiting the oxidation of thiols. Carbon dioxide reacts with peroxynitrite to form the nitrosoperoxycarbonate anion ($ONOOCO_2^-$), which subsequently rearranges to form the nitrocarbonate anion ($O_2NOCO_2^-$). The latter is considered to be the direct oxidant of peroxynitrite-mediated reagents in biological environments [43]. Peroxynitrite-mediated tyrosine nitration is also accelerated in the presence of transition metal ions, either in free form (Cu^{2+}, Fe^{3+}, Fe^{2+}) or as complexes involving protoporphyrin IX (heme) or certain chelators - ethylene diamine tetraacetic acid (EDTA) [15,16,44]. Hence, metal ions catalysis plays an important role in the nitration of protein residues in proteins [45].

Role of Protein Tyrosine Nitration in Pathological Conditions

A variety of proteins are nitrated at tyrosine residues both *in vitro* and *in vivo* by the reaction with peroxynitrite (Table 2) [46-49]. These include, for instance, glutamine synthase [50], cdc2 kinase [51], bovine serum albumin (BSA) [52], MnSOD [53], phosphatidylinositol 3-kinase (PI3K) [54], and tyrosine hydroxylase [55]. A proteomic approach has been able to identify more than 40 NT-carrying proteins that become modified as a consequence of inflammatory responses [5]. The formation of 3-NT proteins has been detected histochemically in inflamed or infected tissues (Figures 2 and 3). Synovial fluid from patients with rheumatoid arthritis (RA) was found to contain a higher concentration of NT-carrying immunoglobulin G (IgG) as compared to osteoarthritic (OA) specimens [56,57]. Peroxynitrite reacts with several amino acids such as tyrosine, phenylalanine and histidine that are modified through intermediary secondary species

[58-61]. Protein sulfhydryls [62] and tyrosyl residues are the principal targets of peroxynitrite in proteins [63,64]. Nitration is maximal at physiological pH (pH 7.4), and its yield decreases under more acidic or basic conditions [64,65].

Protein	Reactive nitrogen Species	Pathological condition	References
ATP synthase complex I (FoF1 complex, F1-ATPase)	Peroxynitrite	inflammation, ischemia	Souza et al (2008), Yeo et al (2008), Tsikas (2012), Peluffo and Radi (2007), Lee et al (2009)
Creatine kinase	Peroxynitrite	myocardial infarction	
Cytochrome c	Peroxynitrite	inflammation, ischemia	
Glyceraldehy 3 phosphate dehydrogenase (GAPDH)	Peroxynitrite	cardiovascular, neurological diseases	
Histone protein (H2A)	Peroxynitrite	systemic lupus erythematosus, rheumatoid arthritis	
Manganese superoxide dismutase (MnSOD)	Peroxynitrite	neurodegenerative diseases	
NADH dehydrogenase (NADH ubiquinone oxidoreductase, Complex I)	Peroxynitrite	ischemia reperfusion	
Sarcoplasmic reticulum Ca^{2+} ATPase (SERCA2, calcium pump 2)	Peroxynitrite	myocardial infarction	
Serum proteins	Nitric oxide	systemic lupus erythematosus	
Succinate dehydrogenase (complex II, succinate-coenzyme Q reductase)	Peroxynitrite	ischemia	
Synovial tissue proteins	nitric oxide	rheumatoid arthritis, osteoarthritis	
Tau protein	Peroxynitrite	Alzheimer's disease	

Table 2: Role of protein tyrosine nitration in various pathological conditions.

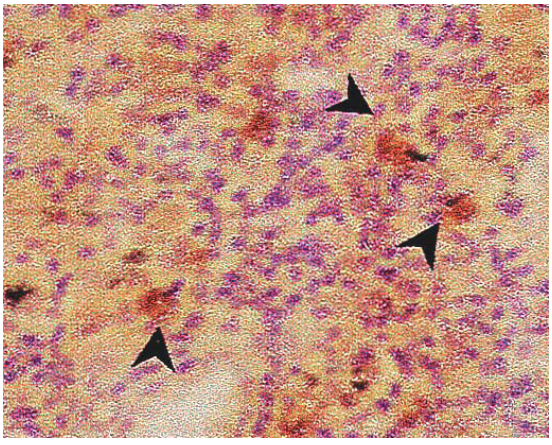


Figure 2: Immunohistochemistry of protein tyrosine nitration. The figure shows nitrotyrosine staining of a human lymph node undergoing nonspecific immune activation. The arrowheads indicate 3-nitrotyrosine immunoreactivity in macrophages (*adapted from Radi et al, 2001*).

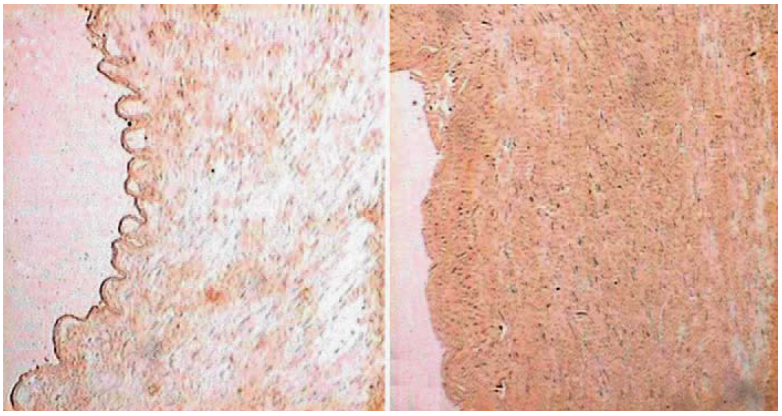


Figure 3: Arteries immunostained with anti-nitrotyrosine antibodies. (A) The nitrotyrosine immunostaining of arteries from a healthy donor (B) chronic kidney disease patient (*adapted from Guilgen et al, 2011*).

Nitration of tyrosine residues can profoundly alter protein structure and function, suggesting that protein nitration may be fundamentally related to and predictive of oxidative cell injury. The subsequent release of altered proteins may enable them to act as antigens inducing antibodies against self-proteins. The biological significance of tyrosine nitration supports the formation of 3-NT *in vivo* in diverse pathologic conditions and 3-NT is thought to be a relatively specific marker of oxidative damage mediated by peroxynitrite and other nitrogen free radical species. The immunoreactivity of 3-NT has been reported in several human pathological conditions. Free/protein-bound tyrosine are attacked by various RNS, including peroxynitrite, to form free/protein-bound 3-NT, which may provide insight into the mechanism of severe disease activity as seen in lupus patients [66]. In addition, numerous other disease states using non-human models have been shown to involve the formation of 3-NT [67].

Elevated levels of 3-nitrotyrosine have been reported is various human pathologies such as atherosclerosis, multiple sclerosis,

Alzheimer's disease, Parkinson's disease and its animals models, cystic fibrosis, asthma, lung diseases, myocardial malfunction, stroke, amyotrophic lateral sclerosis (ALS), chronic hepatitis, cirrhosis, diabetes, etc. [68,69]. Formation of 3-nitrotyrosine in proteins is considered to be a post-translational modification with important pathophysiological consequences and is one of the early markers of nitrosative stress that have been revealed in multiple sclerosis. Elevated contents of nitrates, nitrites, and free nitrotyrosine were found in the cerebrospinal fluid of subjects and have been proposed as functional biomarkers of neurodegeneration [70]. Approximately 1 to 10 residues of tyrosine per 100,000 (10–100 μmol 3-NT/mol of tyrosine) are found to be nitrated in plasma proteins under inflammatory conditions such as in cardiovascular disease [71], although up to 10 times more 3-NT can be detected in tissues [19]. Studies have found a strong correlation between 3-NT plasma levels and coronary artery disease (CAD), identifying the elevated levels of 3-NT as an emerging cardiovascular risk factor [72] (Table 2).

Among the many technologies available, the most efficient and dependable approach for the quantification of 3-NT are gas chromatography-mass spectrometry (GC–MS/MS) and liquid chromatography-mass spectrometry (LC–MS/MS) (Table 3). GC–MS/MS and LC–MS/MS based methods reveal that the concentration of 3-NT in human plasma is on the threshold of the pico mole (pM) to nano mole (nM) range and changes only very little upon disease or intervention. These important findings are suitable to serve as the gold standard and as a measure to test the reliability of alternative techniques, such as GC–MS, high performance liquid chromatography (HPLC) with electrochemical detection, or immunological assays. The various antibody assays also need to be validated by these GC–MS/MS or LC–MS/MS methods [73].

Protein	Nitrating agent	Technique or Method*	References
Bovine serum albumin (BSA)	Tetranitromethane	MALDI, MS	Tsikas (2012), Radi et al (2001), Abello et al (2009), Duncan (2003).
Creatine Kinase (CK)	Peroxynitrite	MS	
Cytochrome c (cyt c)	Peroxynitrite	MALDI-TOF	
Cytochromes P450s (CYP2B6, CYP2E1)	Peroxynitrite	MS	
Hemoglobin (Hb), human	Nitrite/hydrogen peroxide	MS	
Human Serum albumin (HSA)	Peroxynitrite, Tetranitromethane	HPLC, MS, MALDI-TOF	
Low density lipoprotein (LDL) apoB100	Peroxynitrite	MS	
Phosphorylase b	Peroxynitrite	MS	

**MALDI, Matrix-Assisted Laser Desorption Ionization; MS, Mass Spectroscopy; HPLC, High Performance Liquid Chromatography, MALDI-TOF, Matrix-assisted laser desorption/ionisation-time of flight mass spectroscopy.*

Table 3: Measurement of tyrosine nitrated proteins by different methods.

In tissue-based proteomics of diseases, more than 100 nitrated proteins have been identified using mass spectrometric methods or Western blot analysis following fractionation of samples with 2- dimensional (2D) Polyacrylamide gel electrophoresis (PAGE) or immunoprecipitation from tissues, such as brains afflicted with neurodegenerative diseases or atherosclerotic blood vessels. Protein nitration occurs in various diseases with biological selectivity and site specificity [3]. There are many reports on nitrated proteins in various diseases, identified by current proteomic methods such as 2D-PAGE or LC-MS/MS [3,12]. The biological functions and the subcellular locations of identified nitrated proteins are classified into multiple categories. Of the identified nitrated proteins in *in vivo* disease models, 25%, 20%, 28% are derived from mitochondria, extracellular, and cytoplasm or intracellular, respectively. Since, various active redox reactions occur in mitochondria, it is expected to be the center of nitration. The major nitrated proteins identified are shown to be involved in energy metabolism (20%), which includes many redox reactions. Among the nitrated proteins, the active roles of tau, α-syn and LDL have extensively been studied in pathogenesis of Alzheimer's disease, Parkinson's disease and atherosclerosis, respectively [74].

Therefore, the presence of 3-NT in biological samples indicates that reactive NO-derived species are produced *in vivo*, leading to various pathophysiological conditions [67,75,76]. Nitric oxide is known to participate as a cytotoxic effector molecule or a pathogenic mediator when overexpressed by either inflammatory stimuli-induced nitric oxide synthase (iNOS) or over stimulation of the constitutive forms of NOS (eNOS). The autologous proteins may become immunogenic if they are structurally modified post-translationally under physiological and pathological conditions. These chemical modifications reactions include transglutamination, deamidation, glycosylation, oxidation, nitration and proteolytic cleavage. The consequence of these protein modifications may be generation or unmasking of new antigenic epitopes, which will stimulate relevant B cells and/or T cells, thus leading to the breakdown or bypass of tolerance. The protein tyrosine nitration is widely recognized as a hallmark of inflammation that is associated with the up-regulation of iNOS and is not affected by exogenous sources of nitrate/nitrite (NO_3/NO_2) or serum thiols [77-80].

Varieties of post-translationally modified nitrated proteins have been shown to accumulate in apoptotic or inflamed tissues [81]. Hence, the accumulation of nitrotyrosine-containing proteins in tissues that appear as foreign to the immune system might induce an autoimmune response and sustain a chronic inflammatory reaction [82]. Elevated levels of anti-nitrotyrosine antibodies have been measured in the synovial fluid of patients with RA and OA [76], serum of patients with systemic lupus erythematosus (SLE) [83] and after acute lung injury [84]. It has been suggested that alteration in amino acid structure or sequence may generate neoepitopes on self-proteins, leading to an immune attack. Furthermore, autologous proteins may also become immunogenic if they are structurally modified. The modifications may generate or mask antigenic epitopes and stimulate relevant B cells and/or T cells, leading to breakdown or bypass of tolerance [77]. Approaches directed at inhibiting the oxidative modification of proteins caused by RNS may be useful in understanding the pathology of inflammatory, neurodegenerative and autoimmune diseases [85].

Acknowledgment

The author wishes to thank co-researchers and co-authors, Dr. Fahim Khan and Dr. Rizwan Ahmad, who have been associated with studies on various aspects of free radical biochemistry.

References

1. Ischiropoulos H (1998) Biological tyrosine nitration: a pathophysiological function of nitric oxide and reactive oxygen species. Arch Biochem Biophys 356: 1-11.

2. Souza JM, Daikhin E, Yudkoff M, Raman CS, Ischiropoulos H (1999) Factors determining the selectivity of protein tyrosine nitration. Arch Biochem Biophys 371: 169-78.

3. Ischiropoulos H (2003) Biological selectivity and functional aspects of protein tyrosine nitration. Biochem Biophys Res Commun 305: 776-83.

4. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-32.

5. Abello N, Kerstjens HA, Postma DS, Bischoff R (2009) Protein tyrosine nitration: selectivity, physicochemical and biological consequences, denitration, and proteomics methods for the identification of tyrosine-nitrated proteins. J Proteome Res 8: 3222-3238.

6. Clementi C, Carloni P, Maritan A (1999) Protein design is a key factor for subunit-subunit association. Proc Natl Acad Sci U.S.A 96: 9616-9621.

7. Jiao K, Mandapati S, Skipper PL, Tannenbaum SR, Wishnok JS (2001) Site-selective nitration of tyrosine in human serum albumin by peroxynitrite. Anal Biochem 293: 43-52.

8. Greis KD, Zhu S, Matalon S (1996) Identification of nitration sites on surfactant protein A by tandem electrospray mass spectrometry. Arch Biochem Biophys 335: 396-402.

9. Batthyany C, Souza JM, Duran R, Cassina A, Cervenansky C, Radi R (2005) Time course and site(s) of cytochrome c tyrosine nitration by peroxynitrite. Biochemistry 44: 8038-8046.

10. Lennon CW, Cox HD, Hennelly SP, Chelmo SJ, McGuirl MA (2007) Probing structural differences in prion protein isoforms by tyrosine nitration. Biochemistry 46: 4850-4860.

11. Yamaguchi Y, Nasu F, Harada A, Kunitomo M (2007) Oxidants in the gas phase of cigarette smoke pass through the lung alveolar wall and raise systemic oxidative stress. J Pharmacol Sci 103: 275-82.

12. Yeo WS, Lee SJ, Lee JR, Kim KP (2008) Nitrosative protein tyrosine modifications: biochemistry and functional significance. BMB Rep 41: 194-203.

13. Beckman JS, Beckman TW, Chen J, Marshall PA, Freeman BA (1990) Apparent hydroxyl radical production by peroxynitrite: implications for endothelial injury from nitric oxide and superoxide. Proc Natl Acad Sci USA 87: 1620-1624.

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

15. Ischiropoulos H, Zhu L, Chen J, Tsai M, Martin JC, et al. (1992) Peroxynitrite-mediated tyrosine nitration catalyzed by superoxide dismutase. Arch Biochem Biophys 298: 431-437.

16. Zhan X, Desiderio DM (2006) Nitroproteins from a human pituitary adenoma tissue discovered with a nitrotyrosine affinity column and tandem mass spectrometry. Anal Biochem 354: 279-289.

17. Quint P, Reutzel R, Mikulski R, McKenna R, Silverman DN (2006) Crystal structure of nitrated human manganese superoxide dismutase: mechanism of inactivation. Free Radic Biol Med 40: 453-8.

18. Yamakura F, Taka H, Fujimura T, Murayama K (1998) Inactivation of human manganese-superoxide dismutase by peroxynitrite is caused by exclusive nitration of tyrosine 34 to 3-nitrotyrosine. J Biol Chem 273: 14085-14089.

19. Brennan ML, Wu W, Fu X, Shen Z, Song W, et al. (2002) A tale of two controversies: defining both the role of peroxidases in nitrotyrosine formation *in vivo* using eosinophil peroxidase and myeloperoxidase-deficient mice, and the nature of peroxidase-generated reactive nitrogen species. J Biol Chem 277: 17415-17427.

20. Klebanoff SJ (1993) Reactive nitrogen intermediates and antimicrobial activity: role of nitrite. Free Radic Biol Med 14: 351-360.

21. Shao B, Bergt C, Fu X, Green P, Voss JC (2005) Tyrosine 192 in apolipoprotein A-I is the major site of nitration and chlorination by myeloperoxidase, but only chlorination markedly impairs ABCA1-dependent cholesterol transport. J Biol Chem 280: 5983-5993.

22. Thomas DD, Espey MG, Vitek MP, Miranda KM, Wink DA (2002) Protein nitration is mediated by heme and free metals through Fenton-type chemistry: an alternative to the NO/O2- reaction. Proc Natl Acad Sci U.S.A 99: 1269-76.

23. Bian K, Gao Z, Weisbrodt N, Murad F (2003) The nature of heme/iron-induced protein tyrosine nitration. Proc Natl Acad Sci U.S.A 100: 5712-5217.

24. Roncone R, Barbieri M, Monzani E, Casella L (2006) Reactive nitrogen species generated by heme proteins: Mechanism of formation and targets. Coord Chem Rev 250: 1286-1293.

25. Halliwell B (1997) What nitrates tyrosine? Is nitrotyrosine specific as a biomarker of peroxynitrite formation *in vivo*. FEBS Lett 411: 157-60.

26. Radi R (2004) Nitric oxide, oxidants, and protein tyrosine nitration. Proc Natl Acad Sci U.S.A 101(12): 4003-4008.

27. Peluffo G, Radi R (2007) Biochemistry of protein tyrosine nitration in cardiovascular pathology. Cardiovasc Res 75: 291-302.

28. Schopfer FJ, Baker PR, Freeman BA (2003) NO-dependent protein nitration: a cell signaling event or an oxidative inflammatory response? Trends Biochem Sci 28: 646-54.

29. Grzelak A, Balcerczyk A, Mateja A, Bartosz G (2001) Hemoglobin can nitrate itself and other proteins. Biochim Biophys Acta 1528: 97-100.

30. Kilinc K, Kilinc A, Wolf RE, Grisham MB (2001) Myoglobin-mediated tyrosine nitration: no need for peroxynitrite. Biochem Biophys Res Commun 285: 273-276.

31. Lu N, Zhang Y, Gao Z (2009) Nitrite-glucose-glucose oxidase system directly induces rat heart homogenate oxidation and tyrosine nitration: effects of some flavonoids. Toxicol *In vitro* 23: 627-33.

32. Ceriello A, Mercuri F, Quagliaro L, Assaloni R, Motz E, et al. (2001) Detection of nitrotyrosine in diabtetic plasma: evidence of oxidative stress. Diabetologia 44: 834.

33. Wang N, Lia D, Lu NH, Yi L, Huang XW, et al. (2010) Peroxynitrite and hemoglobin-mediated nitrative/oxidative modification of human plasma protein: effects of some flavonoids. J Asian Nat Prod Res 12: 257-264.

34. Witherick J, Wilkins A, Scolding N (2011) Mechanisms of oxidative damage in multiple sclerosis and a cell therapy approach to treatment. Autoimmune Dis 2011: 164608.

35. Frischer JM, Bramow S, Dal-Bianco A, Lucchinetti CF, Rauschka H, et al. (2009) The relation between inflammation and neurodegeneration in multiple sclerosis brains. Brain 132: 1175-89.

36. Negre-Salvayre A, Coatrieux C, Ingueneau C, Salvayre R (2008) Advanced lipid peroxidation end products in oxidative damage to proteins. Potential role in diseases and therapeutic prospects for the inhibitors. British J Pharmacol 153: 6-20.

37. Gonsette R (2008) Oxidative stress and excitotoxicity: a therapeutic issue in multiple sclerosis? Mult Scler 14: 22-34.

38. Dietrich-Muszalska A, Olas B, Głowacki R, Bald E (2009) Oxidative/nitrative modifications of plasma proteins and thiols from patients with schizophrenia. Neuropsychobiology 59: 1-7.

39. Denicola A, Freeman BA, Trujillo M, Radi R (1996) Peroxynitrite reaction with carbon dioxide/bicarbonate: kinetics and influence on peroxynitrite-mediated oxidations. Arch Biochem Biophys 333: 49-58.

40. Gow A, Duran D, Thom SR, Ischiropoulos H (1996) Carbon dioxide enhancement of peroxynitrite-mediated protein tyrosine nitration. Arch Biochem Biophys 333: 42-48.

41. Lymar SV, Hurst JK (1996) Carbon dioxide: physiological catalyst for peroxynitrite-mediated cellular damage or cellular protectant. Chem Res Toxicol 9: 845-50.

42. Lymar SV, Jiang Q, Hurst JK (1996) Mechanism of carbon dioxide catalyzed oxidation of tyrosine by peroxynitrite. Biochemistry 35: 7855-7861.

43. Uppu RM, Squadrito GL, Pryor WA (1996) Acceleration of peroxynitrite oxidations by carbon dioxide. Arch Biochem Biophys 327: 335-43.

44. Beckman JS, Ischiropoulos H, Zhu L, vander Woerd M, Smith C, et al. (1992) Kinetics of superoxide dismutase- and iron-catalyzed nitration of phenolics by peroxynitrite. Arch Biochem Biophys 298: 438-445.

45. Daiber A, Bachschmid M, Beckman JS, Munzel T, Ullrich V (2004) The impact of metal catalysis on protein tyrosine nitration by peroxynitrite. Biochem Biophys Res Commun 317: 873-81.

46. Ohmori H, Kanayama N (2003) Mechanisms leading to autoantibody production: link between inflammation and autoimmunity. Curr Drug Targets Inflamm Allergy 2: 232-241.

47. Aulak KS, Miyagi M, Yan L, West KA, Massillon D, et al. (2001) Proteomic method identifies proteins nitrated *in vivo* during inflammatory challenge. Proc Natl Acad Sci USA 98: 12056-12061.

48. Sampson JB, Ye Y, Rosen H, Beckman JS (1998) Myeloperoxidase and horseradish peroxidase catalyze tyrosine nitrationin proteins from nitrite and hydrogen peroxide. Arch Biochem Biophys 356: 207-213.

49. Pfeiffer S, Lass A, Schmidt K, Mayer B (2001) Protein tyrosine nitration in cytokine-activated murine macrophages. J Biol Chem 276: 34051-34058.

50. Berlett BS, Friguet B, Yim MB, Chock PB, Stadtman ER (1996) Peroxynitrite-mediated nitration of tyrosine residues in *Escherichia coli* glutamine synthetase mimics adenylylation: relevance to signal transduction. Proc Natl Acad Sci U.S.A 93: 1776-1780.

51. Kong SK, Yim MB, Stadtman ER, Chock PB (1996) Peroxynitrite disables the tyrosine phosphorylation regulatory mechanism: lymphocytespecific tyrosine kinase fails to phosphorylate nitrated cdc2(6-20)NH2 peptide. Proc Natl Acad Sci U.S.A 93: 3377-3382.

52. Kamisaki Y, Wada K, Bian K, Balabanli B, Davis K, et al. (1998) An activity in rat tissues that modifies nitrotyrosine-containing proteins. Proc Natl Acad Sci U.S.A 95: 11584-11589.

53. MacMillan-Crow LA, Thompson JA (1999) Tyrosine modifications and inactivation of active site manganese superoxide dismutase mutant (Y34F) by peroxynitrite. Arch Biochem Biophys 366: 82-88.

54. Hellberg CB, Boggs SE, Lapetina EG (1998) Phosphatidylinositol 3-kinase is a target for protein tyrosine nitration. Biochem Biophys Res Commun 252: 313-317.

55. Ara J, Przedborski S, Naini AB, Jackson-Lewis V, Trifiletti RR, et al. (1998) Inactivation of tyrosine hydroxylase by nitration following exposure to peroxynitrite and methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). Proc Natl Acad Sci U.S.A 95: 7659-7663.

56. Evans TJ, Buttery LDK, Carpenter A, Springall DR, Polak JM, et al. (1996) Cytokine-treated human neutrophils contain inducible nitric oxide synthase that produces nitration of ingested bacteria. Proc Natl Acad Sci U.S.A 93: 9553-9558.

57. Uesugi M, Yoshida K, Jasin HE (2000) Inflammatory properties of IgG modified by oxygen radicals and peroxynitrite. J Immunol 165: 6532-6537.

58. Alvarez B, Ferrer-Sueta G, Freeman BA, Radi R (1999) Kinetics of peroxynitrite reaction with amino acids and human serum albumin. J Biol Chem 274: 842-48.

59. Alvarez B, Radi R (2003) Peroxynitrite reactivity with amino acids and proteins. Amino Acids 25: 295-311.

60. Alvarez B, Rubbo H, Kirk M, Barnes S, Freeman BA, et al. (1996) Peroxynitrite-dependent tryptophan nitration. Chem Res Toxicol 9: 390-396.

61. Ischiropoulos H, al-Mehdi AB (1995) Peroxynitrite-mediated oxidative protein modifications. FEBS Lett 364: 279-82.

62. Radi R, Beckman JS, Bush KM, Freeman BA (1991) Peroxynitrite oxidation of sulfhydryls. The cytotoxic potential of superoxide and nitric oxide. J Biol Chem 266: 4244-4250.

63. Beckman JS (1996) Oxidative damage and tyrosine nitration from peroxynitrite. Chem Res Toxicol 9: 836-844.

64. Vander Vliet A, Eiserich JP, O'Neill CA, Halliwell B, Cross CE (1995) Tyrosine modification by reactive nitrogen species: a closer look. Arch Biochem Biophys 319: 341-349.

65. Beckman JS, Chen J, Ischiropoulos H, Crow JP (1994) Oxidative chemistry of peroxynitrite. Methods Enzymol 233: 229-240.

66. Losman MJ, Fasy TM, Novick KE, Monestier M (1992) Monoclonal autoantibodies to subnucleosomes from a MRL/MP-+/+ mouse.

Oligoclonality of the antibody response and recognition of a determinant composed of histones H2A, H2B and DNA. J Immunol 48: 1561-1569.

67. Crow JP (1999) Measurement and significance of free and protein-bound 3-nitrotyrosine, 3-chlorotyrosine, and free 3-nitro-4-hydroxyphenylacetic acid in biologic samples: A high performance liquid chromatography method using electrochemical detection. Methods Enzymol 301: 151-160.

68. Duncan MW (2003) A review of approaches to the analysis of 3-nitrotyrosine. Amino Acids 25: 351-361.

69. Yang B, Rizzo V (2007) TNF-α potentiates protein tyrosine nitration through activation of NADPH oxidase and eNOS localized in membrane rafts and caveolae of bovine aortic endothelial cells. Am J Physiol Heart Circ Physiol 292: 954-962.

70. Calabrese V, Scapagnini G, Ravagna A, Bella R, Foresti R, et al. (2002) Nitric oxide synthetase is present in the cerebrospinal fluid of patients with active multiple sclerosis and is associated with increases in cerebrospinal fluid protein nitrotyrosine and S-nitrosothiols and changes in glutathione levels. J Neurosci Res 70: 580-587.

71. Shishehbor Mviles RJ, Brennan ML, Fu X, Goormastic M, Pearce GL, et al. (2003) Association of nitrotyrosine levels with cardiovascular disease and modulation by statin therapy. JAMA 289: 1675-1680.

72. Souza JM, Peluffo G, Radi R (2008) Protein tyrosine nitration-Functional alteration or just a biomarker? Free Radic Biol Med 45: 357-366.

73. Tsikas D (2012) Analytical methods for 3-nitrotyrosine quantification in biological samples: the unique role of tandem mass spectrometry. Amino Acids 42: 45-63.

74. Lee JR, Kim JK, Lee SJ, Kim KP (2009) Role of tyrosine nitration in neurodegenerative diseases and atherosclerosis. Arch Pharm Res 32: 1109-1118.

75. Ohshima H, Friesen M, Brouet I, Bartsch H (1999) Nitrotyrosine as new marker for endogenous nitrosation and nitration of proteins. Food Chem Toxicol 28: 647-52.

76. Dixit K, Khan MA, Sharma YD, Moinuddin, Alam K (2011) Peroxynitrite-induced modification of H2A histone presents epitopes which are strongly bound by human anti-DNA autoantibodies: role of peroxynitrite-modified-H2A in SLE induction and progression. Hum Immunol 72: 219-225.

77. Ohmori H, Oka M, Nishikawa Y, Shigemitsu H, Takeuchi M, et al. (2005) Immunogenicity of autologous IgG bearing the inflammation-associated marker 3-nitrotyrosine. Immunol Lett 96: 47-54.

78. Khan F, Siddiqui AA (2006) Prevalence of anti-3-nitrotyrosine antibodies in the joint synovial fluid of patients with rheumatoid arthritis, osteoarthritis and systemic lupus erythematosus. Clin Chim Acta 370: 100-107.

79. Anderton SM (2004) Post-translational modifications of self antigens: implications for autoimmunity. Curr Opin Immunol 16: 753-758.

80. Oates JC, Christensen EF, Reilly CM, Self SE, Gilkeson GS (1999) Prospective measure of serum 3-nitrotyrosine levels in systemic lupus erythematosus: correlation with disease activity. Proc Assoc Am Physicians 111: 611-621.

81. Ohmori H, Kanayama N (2005) Immunogenicity of an inflammation associated product, tyrosine nitrated self-proteins. Autoimmun Rev 4: 224-229.

82. Birnboim HC, Lemay AM, Lam DK, Goldstein R, Webb JR (2003) Cutting edge: MHC class II-restricted peptides containing the inflammation-associated marker 3-nitrotyrosine evade central tolerance and elicit a robust cell-mediated immune response. J Immunol 171: 528-532.

83. Khan F, Ali R (2006) Antibodies against nitric oxide damaged poly L-tyrosine and 3-nitrotyrosine levels in systemic lupus erythematosus. J Biochem Mol Biol 39: 189-196.

84. Thomson L, Christie J, Vadseth C, Lanken PN, Fu X, et al. (2007) Identification of immunoglobulins that recognize 3-nitrotyrosine in patients with acute lung injury after major trauma. Am J Respir Cell Mol Biol 36: 152-157.

85. Ahmad R, Ahsan H (2012) Role of peroxynitrite modified biomolecules in the etiopathogenesis of systemic lupus erythematosus. Clin Exp Med (*Epub ahead of print*).

86. Radi R, Peluffo G, Alvarez MN, Naviliat M, Cayota A (2001) Unraveling peroxynitrite formation in biological systems. Free Radic Biol Med 30: 463-488.

87. Guilgen G, Werneck ML, de Noronha L, Martins AP, Varela AM, et al. (2011) Increased calcification and protein nitration in arteries of chronic kidney disease patients. Blood Purif 32: 296-302.

Advances in Protein Chemistry

Chapter: The Recent Methodologies of Protein Structural Determination and Dynamic Developments

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published November, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

The Recent Methodologies of Protein Structural Determination and Dynamic Developments

Mahmood Rasool[1]*, Arif Malik[2], Abdul Manan[2], Aamer M Qazi[3], Mahmood H Qazi[4]

[1]Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia
[2]Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore, Pakistan
[3]Cancer Genomics, Ontario Institute for Cancer Research, Toronto, Ontario, Canada
[4]Center for Research in Molecular Medicine (CRiMM), The University of Lahore, Pakistan

***Corresponding author:** Dr. Mahmood Rasool, PhD, Assistant Professor, Center of Excellence in Genomic Medicine Research, Post Box No. 80216, King Abdul-Aziz University, Jeddah 21589, Saudi Arabia, Tell: +966-582-254267; E-mail: mahmoodrasool@yahoo.com

Introduction

Various functions of biological system depend upon the structure and function of proteins. Determination of structure and functions of proteins assist in scrutinizing the dynamics of proteins. The study of dynamics in respect to proteins gives insight into protein folding, enzyme function, and target identification ligand binding to the receptor, interaction of protein with other molecules including protein itself, miscoding and/or misfolding of proteins associated with diseases. There are many available methods for the analysis of protein dynamics in which nuclear magnetic resonance (NMR) and x-ray crystallographic methodologies are of most powerful one. These techniques provide excellent high resolution structures of biological molecules including proteins. These help in determination of mean position of atomic particles in molecules and/or crystal and show their mean-square displacements from those positions. More than 100 different structures of various proteins have been studied by X-ray diffraction analysis. Polypeptide chain is the basic component of these structures, consisting of sequence of amino acid residues. Twenty commonly occurring structural units (amino acids) are responsible for the synthesis of polypeptide chains, and in some cases a non-protein group forming a part of combined with protein refers as prosthetic group. Polypeptide chain in a globular protein is folded compactly forming a three-dimensional (3D) structure by various intermolecular forces like van der Waal's interactions, as well as hydrogen bonding with water molecules which usually occur near the protein surface [1].

X-ray diffractions analysis is possible to perform on globular protein with bonded ligands (a molecular that selectively binds to another). The interpretation of biological function is not possible without determining the structure, function and dynamics of biological molecules e.g. functioning of enzymes are very important for understanding the biological attributes of biological system. For the understanding of dynamics, physic-chemical aspects are crucial for making the best insight into the cell and its relevant compounds.

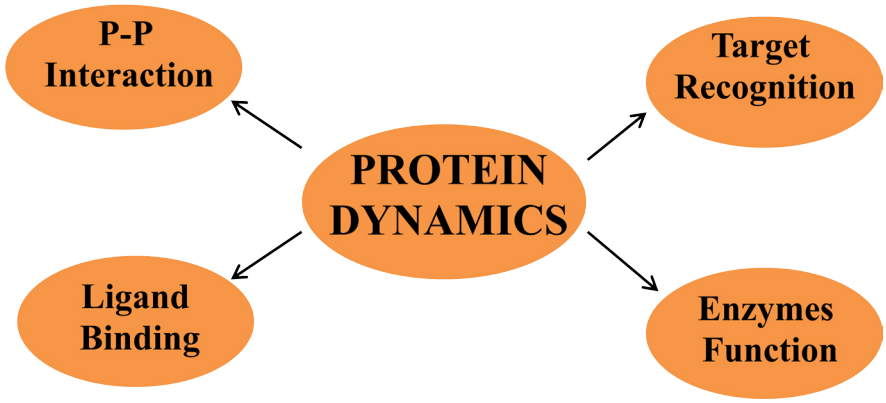


Figure 1: Protein Dynamics Deal With Various Aspects of Proteins.

Protein Folding

The dynamic behavior of proteins is important in order to understand of their function and folding. Proteins are synthesized as linear chains of residues that rapidly and specifically assume their native, folded state. The native state of a protein is usually associated with a compact globular conformation possessing a rigid and highly ordered structure. Proteins are in steady motion. Such motion or dynamics is the cumulative (collective) effect of the forces upon and exerted by, all atoms of amino acids that make up the protein and its surrounding environment. Inside the cell, proteins spontaneously fold into their native conformations under physiological conditions; self-assembly indicate no requirement for an external template to guide folding (as happens in case of transcription). For example: insulin is translated as a single polypeptide precursor (proinsulin). Proinsulin (3 disulfides) folds and refolds. Insulin is post-translational modification (PTM) of proinsulin arising from the proteolytic removal of a 33 residue segment. Local segments fold together and associate with adjacent segments until the entire structure is folded. Most proteins undergo some form of modifications subsequent translation; such modifications are called post-translational modifications (PTMs) including glycosylation, phosphorylation, and sulfation.

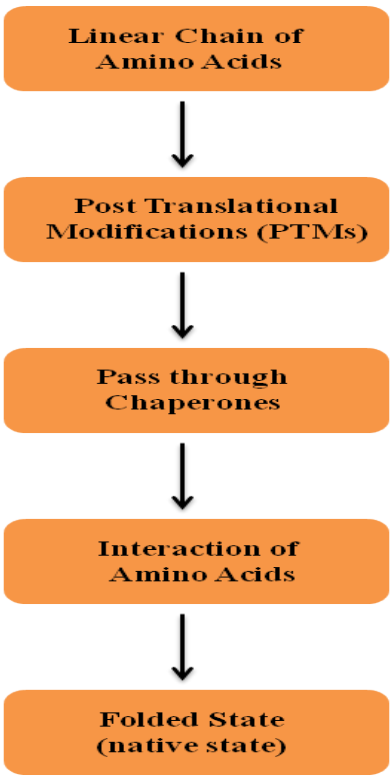


Figure 2: Various Stages of Protein Folding in a Living Cell.

The folding of proteins in the living cell is a complex process and accompanied by at least two crucial factors: The sphere of unfavorable contacts with other molecules in cell.

The appearance of the incorrect intra-molecular contacts at a co-translational folding of amino acids.

A set of special protein-helpers provide assistance for the correct folding to occur, like chaperones and enzymes regulating cis/trans proline isomerization and the formation of the disulfide bridges, that avert protein aggregation and misfolding, accelerate folding, and take part in protein transport (e.g., in protein translocation through the membranes).

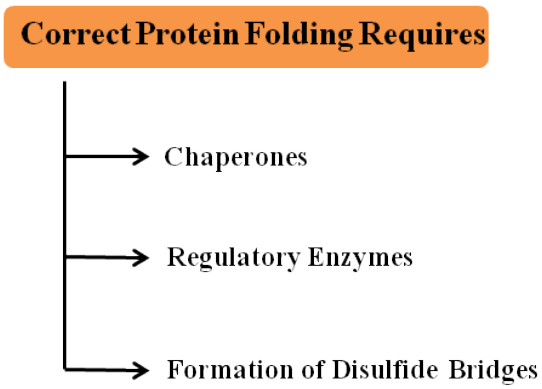


Figure 3: Correct Folding of Proteins in Living Cell Requires Basic Three Components.

Dynamics of Cell Regulation

Dynamic regulation of cell function is controlled in large part by transforming protein activities; though, the diversity of specific tactics by which this is achieved is amazing. The functional diversity of proteins, mutual with the variability of signals related to the various intra-and intercellular processes handled by these proteins and their capability to produce multi-variant and multi-directional responses allow them to form a unique regulatory net in a cell. Several mechanisms that are responsible for dynamics of cell regulation,

are due to altering protein activities, such as changes in protein or mRNA concentration, protein trafficking and/or retention, and post-translational modifications. Proteins are related to regulation of cell functions during various mechanisms such as ubiquitination in cell signaling [2].

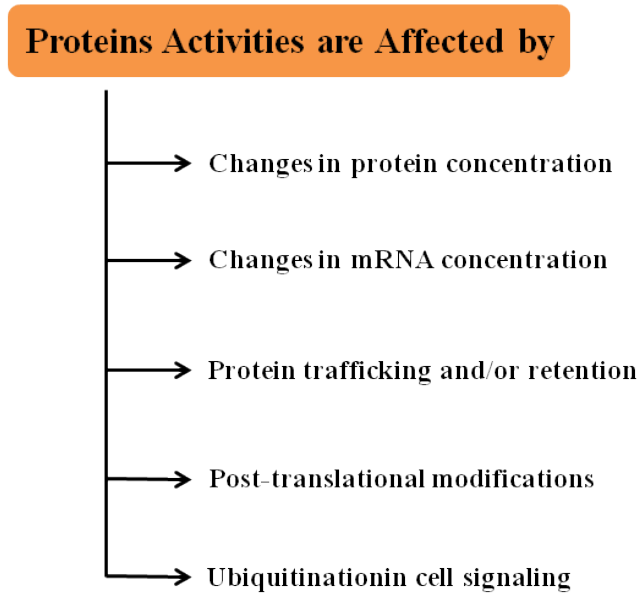


Figure 4: Protein Activities are affected by Multiple Cellular Mechanisms Listed in Diagram.

Endocytosis and cell dynamics

Endocytosis is the process by which the cell membrane of a cell folds inwards to ingest material. The process of endocytosis of cell surface receptors is regulated by caveolae (uncoated vesicles) or by clathrin. Another important protein, called arrestin is responsible for clathrin-mediated endocytosis which was considered as desensitizer for seven transmembrane (7-TM) receptors. Along with the endocytosis, arrestin is also involved in the internalization of other receptors like transforming growth factor beta (TGF-β), produced by skeletal cells. The role of arrestin in endocytosis is regulated by ubiquitination of arrestin, during cell signaling.

Signaling proteins

The several signaling proteins, involve in various cellular mechanisms are dynamic in nature and this dynamic attitude directly plays a role in cell regulation. Cell surface receptors or plasma membrane receptors are internalized throughout the process of endocytosis and signal reduction occurs, and these receptors may be recycled or degraded inside the cell. The control of cell proliferation may experience insult due to defects in the endocytosis of cell surface receptors and may lead to uncontrolled cell division, known as cancer. The process of ubiquitination play a vital role in signaling for endocytosis and degradation of cell surface receptors after the internalization is done. Signaling cascade for endocytosis is regulated by the monoubiquitintion of receptors while polyubiquitination shows degradation of internalized receptors. Moreover, mutation in genes, coding for the process for ubiquitination pathway may produce cancer, e.g. defect/mutation in caveolin-1 has been associate with breast cancer, which is involved in endocytosis.

Protein regulation by Small RNAs

Small RNAs or microRNAs (miRNAs) are responsible for the control of protein levels in the cell. They correspond to 0.5-1.0% of all genes present in organisms including humans. These miRNAs reflect the genes involving other genes regulatory factors like DNA-binding transcription factors. The main role of miRNAs has been associated with development. Moreover, cell regulation in neurons is also linked with miRNAs by controlling levels of protein expression in growth cones and nearby the synaptic membrane. As the miRNAs are involved in protein regulation, similarly protein are involved in cell regulation, called adaptor proteins, which are the part of macromolecular complexes, comprising interaction between protein and lipids as well as the interaction of protein with other proteins. Adaptor proteins are well recognized in immune cells. These included transmembrane adaptors (TRAPs) and cytoplasmic adaptors. Some of these proteins are involved in positive regulation of cell signaling and some play a role as inhibitor in signaling.

Protein Dynamics and Membrane Organization

Cell membrane also shows the movement of proteins present in plasma membrane either in the form of lateral protein of peripheral protein. The movement of protein molecules in plasma membrane is essential for the function and interaction of proteins with respect to other biological molecules. The measurement of movement and interaction with other molecules gives insight into the organization of cell membrane. It is necessary to establish the local distribution of integral membranous protein that is present in specific regions by reducing lateral mobility of proteins in cell membrane. For example, claudins are responsible for making a barrier that inhibits the mixing of proteins in apical and lateral regions of the membrane. Another, important phenomenon of claudins is selective permeability for various solutes. Moreover, it is associated with macromolecular assemblies controlling cell polarity.

Biological membranes are linked with the restriction of cellular compartments. Varieties of different functions by membranous proteins are due to their diversity in structure and showing the feature of hydrophilicity and hydrophobicity. The membranous proteins exhibit hydrophilic region and hydrophobic region, extrinsically and at core surface respectively. Moreover, several membranous proteins are connected with different disorders like brain, lung edema, diabetes etc. In the case of phagocytosis of drug uptake about 70% drugs are delivered through plasma membrane where the drug target is the membranous protein in the form receptor. About 30% f the genes are responsible for encoding the membranous proteins in which receptors are the most important one. As compared to more

than ten thousand unique structures of soluble proteins only few of them have been studied at atomic resolution/level. The complete understanding of the structure and function of membranous proteins is crucial for designing better therapeutical molecules.

Plasma membrane: A complex lipid-protein compound

Plasma membrane consists of lipids with highest percentage (60-70%) followed by the protein with lesser in percentage (30-40%). In a few percentage cholesterol and other molecules are also present in plasma membrane. In the complex structure of plasma membrane the mobility and diffusion of protein become irregular with respect to diffusion coefficient and showing lower value as compared to the artificial membrane systems. Probably, all this happens due to interaction of protein molecules with other molecules, cytoskeleton interaction and presence of sterol and sphingolipid enriched domains known as lipid rafts [3-7].

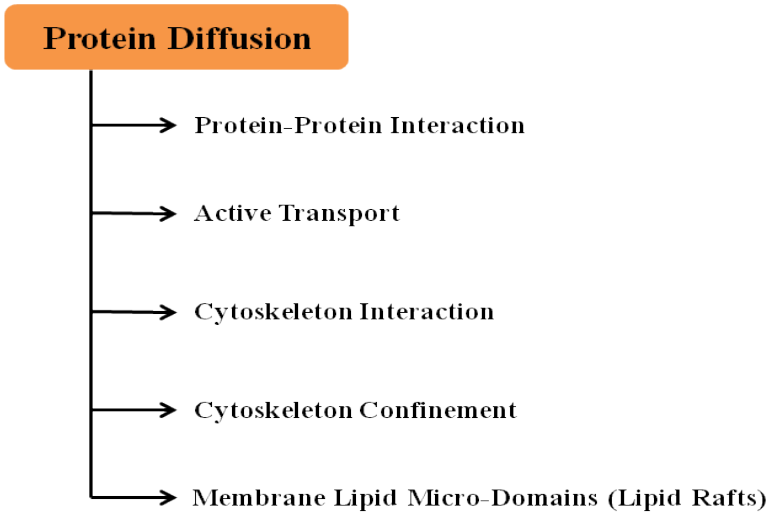


Figure 5: Protein Diffusion in Cell is affected by Multiple Biological Processes.

Particularly three techniques are used for analyzing dynamics in proteins including fluorescence recovery after photo bleaching, fluorescence correlation spectroscopy and single particle tracking. Moreover, physical models are engaged in analyzing the data for protein dynamics that represents extraordinary features of the biophysical nature regarding protein dynamics and membrane domains in membranes. Up till now, there remain considerable unknown information among cholesterol dependent lipid micro-domains, protein interactions, and cause of the underlying cytoskeleton. Innovative techniques of microscopy may provide better sequential and spatial resolution ensuing in more perfect identification and recognition of proteins as well as its dynamics in biological system. The kinetic aspects of protein in a membrane and the types of motion that it contain go through by the lateral organization of cellular boundary. Fluid mosaic model given by Singer and Nicolson projected that the cell membrane contains membrane-associated proteins which can diffuse (spread) liberally with Brownian movement [8] and randomly distributed on the cellular surface.

Importance of Dynamics in Proteins

The dynamic behavior of proteins is distinguished to play significant role in protein function. Diverse characteristics of protein function are influenced by dynamics of protein. For example, protein to protein identification [9], protein-DNA associations [10] and enzyme-substrate bonding and activity of enzymes [11-13] are all resolute by the structural flexibility of the protein molecule and side chains, resulting into characterizing not only the structure, but also dynamic properties as well. On a wide variety of time scales, the dynamics of proteins are thoroughly related to function of protein. Fast as well as moderately fast fluctuations of protein structure facilitate a protein to sample a complex conformational-energy landscape. Such motions of atoms or residues give rise to the slower processes associated with protein function. Molecular dynamics (MD) simulations have shown that a protein can sample thousands of conformations within a very short time. Determination of protein dynamics by X-ray, NMR and other experimental techniques and theory helped in understanding the dynamics that provides an important connection between protein function and protein structure [14].

Dynamics of proteins and mRNA

mRNA and protein molecules are dynamic entities, and their presence in the cell is the outcome of contrasting processes that bring about their biosynthesis or destruction. The abundance-weighted total of all proteins in a cell or sub-cellular space is also dynamic. The amount of a true intracellular protein in a cell is the result of the opposing processes of protein synthesis and protein degradation. However, for extracellular proteins, it has to factor in the irreversible loss due to secretion. If a protein decreases in amount in a cell, this can be a consequence of a reduction of mRNA, or of a decrease in ribosomal activity or translation initiation. Equally possible it could be the consequence of enhanced degradation of the protein.

Protein Motions

Proteins are composed of multiple domains, whose flexibility and mobility lead to a great deal of versatility in their function. Protein dynamics, at the domain level, is a controlling influence in the allosteric formation of protein complexes, in catalysis, in cell signaling and regulation, in cellular locomotion and in metabolic transport.

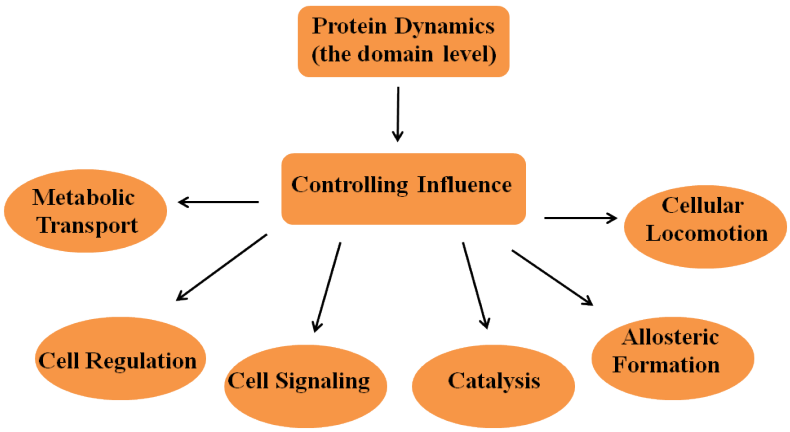


Figure 6: At Domain Level, Protein Dynamics Control Various Processes.

In cells, membrane channels and receptors are often assembled into macromolecular complexes in specialized sub-cellular domains for the dynamic control of diverse cellular events. For instance, forming quaternary complexes of receptors, such as the EGF receptor or the PDGF receptor is necessary for initializing cascades of signaling events for cell growth and proliferation [15].

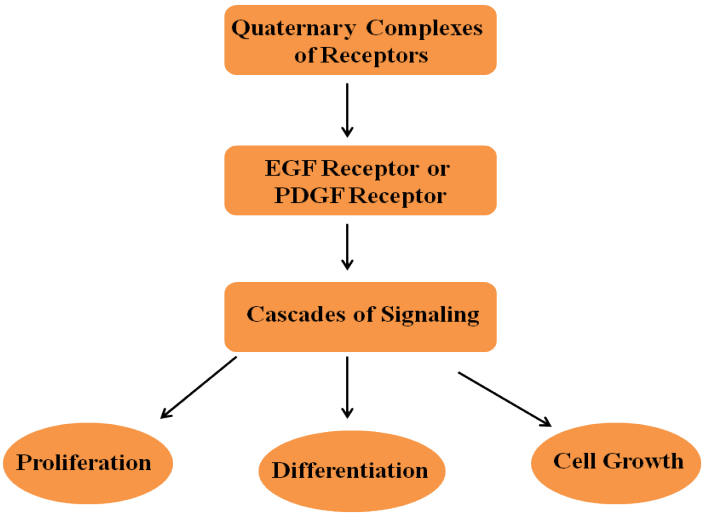


Figure 7: Membrane Channels and Receptors are assembled into Macromolecular Form for the Dynamic Control of Diverse Cellular Events.

The function of ion transport proteins, such as the cystic fibrosis transmembrane conductance regulator (CFTR) or sodium–phosphate co-transporter 2a (NaPiT2a), is regulated by a network of interactions with other membranous proteins like G-protein coupled receptors and other ion channels by forming membrane oligomers either directly, or via cytosolic proteins as adapters or scaffolds. Forming large adherence membrane complexes at the cell-cell junctions is essential to maintain tissue integrity and to suppress tumor cell invasion [16-18]. Understanding how transmembrane protein complexes are regulated and deregulated in disease state can help to identify elements as target to treat various diseases. Protein dynamics arises as a result of interplay between the mechanical forces and the thermal forces (the forces due to the collision of the protein with solvent molecules). These thermal forces are random in magnitude and direction, help protein for a process called as diffusion. A freely diffusing object displays motion, called Brownian motion, with frequent changes in the direction and speed of its movement. Proteins obey Brownian dynamics. The world in which proteins function is characterized by the presence of a significant amount of noise and the resultant diffusion of protein subunits arising from thermal motion. This thermal motion is indispensable for the protein to attain its equilibrium state.

Importance of conformational flexibility

The conformational flexibility of amino acids is a major contributor in the biochemical functionality of proteins (Dodson and Verma, 2006; Teilum et al., 2009). Protein motions are of functional significance having range from fast (sub-nanosecond) atomic level fluctuations to slow (microsecond upward), large-scale amino acid residues conformational reorganization [19] Henzler-Wildman et al., 2007). It has been observed by several studies internal motions of protein is directly related to biochemical functions [20] Smock and Gierasch, 2009) and normal mode analysis helps in the identification of categorization and prediction of large-scale conformational changes in the protein (Ma, 2005) and elastic-network models (Hall et al., 2007; Keskin et al., 2000) has been quite successful too. Molecular Dynamics (MD) simulations are effectively helpful to investigate the conformational energy landscape (Karplus and Kuriyan, 2005; Karplus and McCammon, 2002) and approaching the idea that how protein dynamics relates back to sequence.

Proteins: Flexible nanoparticles

Proteins are flexible nanoparticles that commonly achieve their biological function by collective atomic motions. Flexibility refers that protein molecules are able to change their conformation in the biological system more quickly than any other macromolecule especially in membranous system. Such motions may be differentiated by hinge, shear, or rotational motions of entire protein domains,

loop movements or slight rearrangements of amino acid side chains. In many cases it is far from understandable how collective motions are related to a particular biological task.

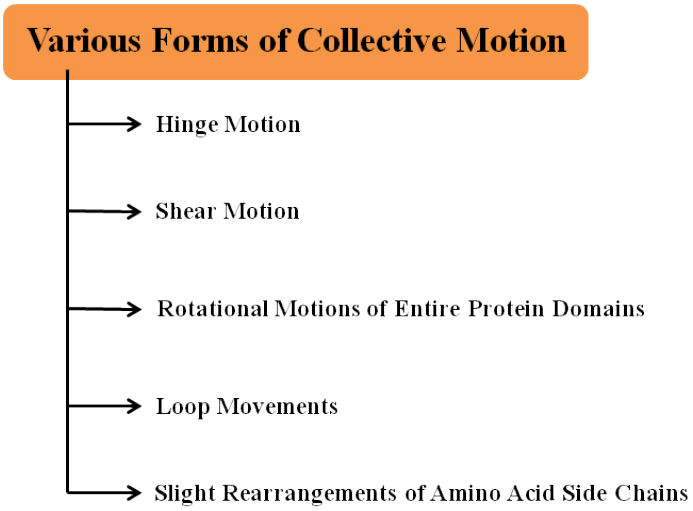


Figure 8: Types of Collective Motions in Proteins.

Myoglobin (MB) is the O2-shuttling protein in animal muscles. Structure of Mb does not possess a static channel for O2. Mb must undergo large-scale structural fluctuations to form a dynamic path so that O2or other gas ligands like CO can move in or out. Thus the gating of such a path involves not only energy (enthalpic) barriers but also a conformation (entropic) limit. This may be one of the reasons why proteins are usually large.

Functional Mode Analysis (FMA)

Proteins regularly bring about their biological function by collective atomic motions. The identification of collective motions related to a specific protein function from a molecular dynamics trajectory is important. A novel technique, Functional Mode Analysis (FMA) is used to detect the collective motion that is directly related to a particular protein function. It is based on an ensemble (a group of structure models) together with random functional quantity that measures the functional state of the protein. This technique detects the collective motion that is maximally correlated to the functional quantity. The functional quantity could be related to a geometric, chemical or electrostatic observable, or any other variable that is significant to the protein function. In addition, the motion that exhibits the largest possibility to induce a considerable change in the functional quantity, is predictable from the given protein ensemble. For the determination of the correlation between the given ensemble and functional quantity of proteins two different measures is applied:

The Pearson correlation coefficient measures linear correlation.

Provides the mutual information that can evaluate any sort of interdependence.

Functional mode analysis (FMA) detect the maximally correlated motion allows to develop a model for the functional state in terms of a single collective coordinate. The new approach is illustrated using a number of biomolecules, including Trp-cage, T4 lysozyme, polyalanine-helix, and leucine-binding protein.

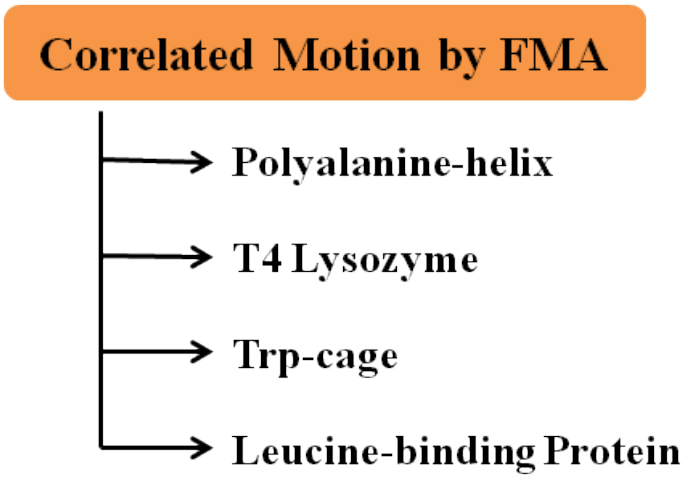


Figure 9: FMA Detect Correlated Motion and Has Been Used for Illustrating a Number of Biomolecules.

Computer simulation offers the possibility to study biomolecules and their dynamics in great aspect, at high spatial and temporal resolution, by this means complementing information that is accessible by experiment [21]. Molecular dynamics (MD) simulations which based on Newtonian mechanics are widely used and well-developed approach to obtain atomic-level resolution information on the dynamics of molecular systems over time, mainly for proteins in aqueous solution [22].

Brownian Dynamics (BD) method

Protein interactions are responsible for most cellular processes. As far as the protein function is concerned, it is necessary to get an idea that how protein interacts with other molecules in the biological system at the molecular level. Computational and experimental approaches could be employed for this purpose. Experimental techniques, like nuclear magnetic resonance (NMR) and X-ray crystallographic analysis are used to explore protein structures that could be accessed through the Protein Data Bank (PDB) on internet. To determine the structures of macromolecules like protein complexes is more difficult by employing these methodologies. The Brownian dynamics (BD) method (Monte Carlo approach), has been used to envisage protein interactions. Brownian dynamics was effectively used to simulate the recognition between scorpion toxins and potassium channels [23]. BD prediction for the interaction between KcsA potassium channel and scorpion toxin Lq2 has been verified by the potassium channel-charybdo toxin complex structure [24], which was determined by NMR studies [25].

Various types of methods for protein structural determination and dynamics can be divided into "Experimental Methods" and "Computational Methods".

Experimental Methods: Techniques for Analyzing Protein Dynamics

The dynamics and of proteins can be explored at atomic or residue level (amino acids), with the help of established methodologies/techniques, like hydrogen exchange NMR or Molecular Dynamics simulations. These techniques have supplied the information associating structure and dynamics but they cannot be applied easily on a proteomic scale and are not able to expose evolutionary linkages clearly. Free energy estimation-based models are constructive to calculate local properties of protein, such as hydrogen exchange rates [26]. The technique requires widespread calculations and assessment at residue (amino acids) level of a thermodynamic quantity and the free energy (ΔG) of folding which is incredibly complex to calculate perfectly even using careful parameterizations. Elastic Network Models (ENM), are very constructive in exposing slow motions of protein molecules. Such models do not provide specific interactions within the molecules therefore; can offer limited approaching into the physicochemical properties of highly dynamic protein. Simple and consistent methods are needed for *in silico* (computational) analysis that could help to recognize and give details the idea for dynamics of proteins.

Techniques for analyzing protein dynamics:

Fluorescence recovery after photobleaching (FRAP)

Single particle tracking (SPT)

Fluorescence correlation spectroscopy (FCS)

Nuclear Magnetic Resonance (NMR)

X-Ray and Neutron Scattering

Fluorescence Technique

Single Molecule Technique

Hydrogen Exchange Mass Spectrometry

Fourier Transform Infrared (FTIR) Spectroscopy

Circular Dichroism (CD)

Raman Spectroscopy

Dual Polarization Interferometry (DPI)

Crystallographic Analysis

Electron Crystallography

Atomic Force Microscopy

Cryoelectron Microscopy

Above mentioned techniques and established physical models are engaged into data analysis and give extraordinary explanations of the biophysical characteristics of protein dynamics and micro-domains in cell membranes. In fluorescence imaging methods and microscope systems are used in green fluorescent protein (GFP) biology that makes it simple to identify the position of GFP fusion proteins, moreover, to quantify their profusion and to investigate the motion and interactions. Imaging methods like FRAP, FRET and FCS have been modified that they could be available on commercial scale as user-friendly scanning microscopes. Moreover, computing resources are available in huge amount and data can be analyzed into digital information with the help of software. The advances in imaging methods and technical equipment are helpful in providing a marvelous insight for exploring the kinetic properties of proteins in biological systems.

Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance (NMR) spectroscopy is employed to scrutinize the dynamic behavior of a protein at a multitude of specific sites. Moreover, protein movements on a broad range of timescales can be screened using various types of NMR experiments, like nuclear spin relaxation rate measurements give insights into the internal motions on fast (sub-nanoseconds) and slow (microseconds, μs to milliseconds, ms) timescales and generally rotational diffusion of the molecule (5-50 nanoseconds, ns), whereas rates of magnetization transfer among protons with different chemical shifts and proton exchange give an idea about movements of protein domains on the very slow timescales (milliseconds to days). X-ray crystallographic analysis and NMR spectroscopic analysis regularly present ideas of the function of protein.

The high resolution NMR spectroscopy is found to obtain comprehensive information about the structure and dynamics of proteins, in order to explicate their functions. In order to function accurately, the polypeptide chain must fold into a well-defined, compact structure (3D structure). In general, non-polar amino acid side chains pack into the hydrophobic interior core of the molecule, while hydrophilic side chains make up the solvent-accessible protein surface. A folded protein is well planned and well ordered and significant portions of the protein are often flexible as well. Both well-ordered and flexible protein domains have roles in biological processes. The 3D structure of a protein is specified by the linear sequence of amino acids in the polypeptide chain. Despite progress in predicting the structure of a protein from its amino acid sequence, experimental methods remain the only reliable means to obtain high-resolution

protein structures. The first protein structures were solved at high resolution over 40 years ago by X-ray diffraction. When a single crystal of proteins was available, this method remains the most efficient means to obtain high-resolution proteins structures. About 15 years ago, it was shown that solution structures of proteins having molecular weights below 10kDa could be solved using 2D homonuclear NMR techniques. The achievement made it possible, for the first time, to view structures of protein in the absence of crystal contacts and cemented the way for NMR studies of larger proteins. The study of larger proteins required the development of 2D/3Dheteronuclear NMR spectroscopy. This approach was made practicable by recombinant DNA technology, which permitted efficient incorporation of C^{13}, N^{15} and H^{2} spins into proteins and by advances in instrumentation.

Applications of NMR

NMR has become a refined and influential analytical technology that has found a variety of applications in many disciplines of scientific research, medicine, and various industries. In concert with X-ray crystallography, NMR spectroscopy is one of the two most important and leading technologies for the structure determination of biological macromolecules at atomic resolution. Applications of NMR spectroscopy are listed below:

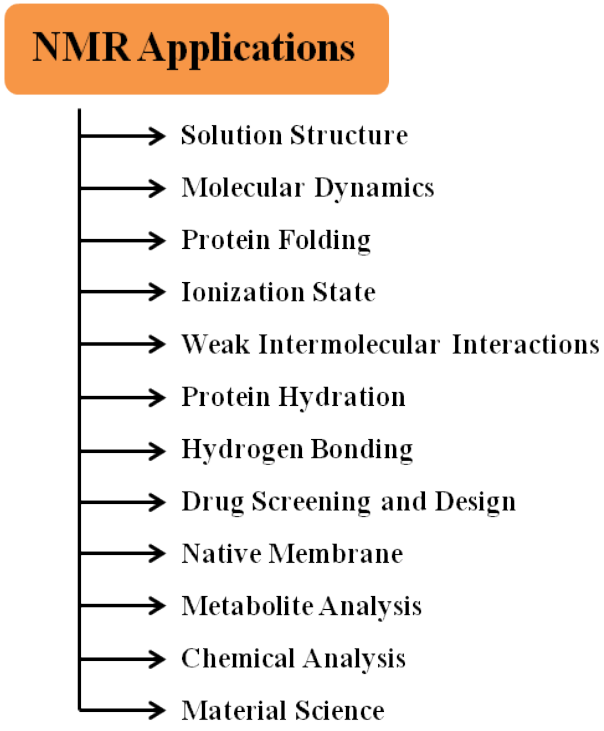


Figure 10: Applications of Nuclear Magnetic Resonance (NMR).

Study of protein internal motion

Character of protein internal motion largely requires local optical probes and unresolved hydrogen exchange as well as one dimensional NMR techniques that could reveal a surprising complexity and richness in the internal motion of proteins. The number of exclusive topological folds that have been characterized at high resolution by crystallographic and nuclear magnetic resonance (NMR) based methods. Nuclear magnetic resonance (NMR) offers many opportunities to the characterization of a variety of dynamic phenomena in proteins at atomic resolution.

NMR relaxation phenomena

The information obtained from NMR dynamics experiments provides insights into specific structural changes or configurationally energetic associated with function including protein-protein interactions, ligand binding, enzyme function and target recognition. Structure identification by NMR is based on the conversion of force-field derived and experimentally determined distances into a set of three dimensional (3D) coordinates. Large local structural differences between generated protein structures can correspond to flexibility of structure but due to the result of a lack in experimental data. Diverse nuclear magnetic resonance (NMR) techniques are employed to acquire information of the motions within proteins at molecular level on several different time-scales. Computer programs are helpful to investigate protein motions, with the help of accessible structural information. Dynamic fluctuations at molecular level have been found as a driving force for multiple types of interactions, in addition to the spatial arrangement of atoms in proteins. Most biological processes in a biological system are tightly regulated, such as immune response, transcription regulation or signaling, in which dynamic contributions have been found as a most important regulatory control in a living cell.

Energy landscape and conformational coordinates

The model of energy landscape illustrates that a protein displays different populations of conformational/structural coordinates. The population of the structures follows the Maxwell-Boltzmann distribution and the energy barrier present between different sub-structures exhibits their rate of inter-conversion. A hypothesis based on explorations is that a certain state of protein, contributes to a specific function which is characterized by its temporary spatial arrangement and this leads to an intricate picture of protein motions referring different timescales and amplitudes. Enzyme activity elaborated in the energy landscape model showing that a protein as flexible scaffolds that is mandatory to change its spatial and structural arrangement at some stage in successful interactions to the ligand or other protein molecule. Both enzymes and protein are dynamic molecules but in structure they are static entities. Structure of proteins obtained by the NMR and X-ray crystallographic analysis provide imperative insights into the activity and function of enzymes [19]. Static structures of

protein demonstrate the lowest energy (ground state) conformation and function that rely on higher energy conformations of the enzyme and its substrates [19].

Protein folding is normally argued with the help of energy landscape, showing folding funnel, which portrays thermodynamic development through structures or sub-states leading towards the ground-state conformation of the protein. Multiple folding paths consist of various combinations of sub-states and all available protein conformations correspond to the conformation ensemble. Various computational and spectroscopic methodologies have been developed to observe multiple aspects of protein dynamics. X-ray crystallographic analysis yields some dynamic information. Temperature factor is susceptible to the mean square displacements of atoms because of thermal motions and can be acquired for nearly all heavy atoms. On the other hand, it does not give details on the time scale of thermal motions and difficulty associated with crystal lattice contacts, refinement and static disorder procedures make their explanation relatively difficult [19].

X-Ray and neutron scattering

Neutron scattering and X-ray give insights on dynamics of protein. X-ray scattering usually illustrates global/large changes in size of protein and shape in a time-resolved manner and neutron scattering illustrates on amplitudes and time scales (10-12 to 10 8 s) for hydrogen atom according to its location in protein structures.

Fluorescence technique

It is an important technique in visualizing the dynamics of protein both in single molecule and in the ensemble. Single molecule experimentations are demonstrating to be helpful, leading towards the understanding of the motions of individual molecules of protein as well as demonstrate how to form through translation into an ensemble signal. Computer program/simulations provide as theoretical ground for predicting and analyzing protein motion, processing inputs from variety of experimental techniques, and probing dynamic information beyond what can be evaluated practically. *In silico* techniques help to understand dynamic data that cannot be provided by experimental methodology. Several computational programs have been developed to obtain the data/information on protein dynamics and structural changes. Among the computational programs the Monte Carlo (MC) and Molecular Dynamics techniques are most famous one. The precision of these approaches rely on the protocols used and on the length of simulation. By using the realistic force fields, at most a few nanoseconds (10^{-9} sec) for a small protein in aqueous environment could be reproduced within acceptable computer time. More than 90% the internal protein motions are due to free and small fluctuations of atoms. Such internal motions show local effect and are taken as unimportant for the analysis of protein function. The Concerted Motions (motions of interest) are those that extend over a large number of atoms and responsible for the large structural alteration in the protein. NMR is an effective experimental technique used for the study of protein dynamics. Time scale available to NMR ranges from 10-12 to 105 sec, and covers all the relevant dynamic motions in proteins.

Hydrogen exchange mass spectrometry

Hydrogen exchange together with mass spectrometry (MS) is used for the study of protein dynamics. The specific protein functions depends upon the dynamics of protein such as protein translocation from one side to the other while binding with ligand other macromolecules or conformational changes during enzyme activation. Hydrogen exchange method is made possible because reactive hydrogen in protein is exchanged with deuterium atoms when protein is placed in heavy water solution and the protein mass is measured with the help of high-resolution mass spectrometry. The position of deuterium assimilation is found by examining deuterium merging in peptic segments that are prepared after the labeling reaction.

Positions of hydrogen at peptide amide linkages (backbone amide hydrogen) are substituted with deuterium within 1-10sec when peptides are incubated in heavy water. In case of folded proteins some backbone amide hydrogen is replaced quickly and other replace after the periods of months. Rates of the majority slowly exchanging hydrogen may be reduced (Englander and Kallenbach, 1984).

Fourier Transform Infrared (FTIR) Spectroscopy

Infrared (IR) spectroscopy is one of the well recognized experimental techniques for the investigation of secondary structure of polypeptides as well as proteins. IR spectrum can be achieved for proteins in a large range of environments with a small amount of sample. IR gives information on protein dynamics and structural stability. Fourier transform infrared (FTIR) spectroscopy is well-established and valuable instrument for the inspection of protein conformation in water (H2O) based solution, as well as in deuterated (D_2O) forms and dried states, ensuing in the scrutinizing the protein secondary structure and protein dynamics. Moreover, FTIR spectroscopy assists in measuring the wavelength and intensity of the absorption of infrared (IR) radiation by a sample. The IR spectral data of sample (especially high polymers) are generally interpreted in terms of the vibrations of a structural repeating unit. The repeating units in polypeptide and protein provide nine characteristic IR absorption bands (amide A, B, and I to VII). Furthermore, the amide I and II bands are taken as two most important vibrational bands of the protein back-bone [27].

Infrared spectroscopy is used for the diagnosis of cancer, due to the sensitivity of the technique to alterations in biochemistry of biological systems which escort pathological stages. Infrared radiation (IR) is absorbed by biological system (cells, tissues and fluids) to promote vibration of the covalent bonds of molecules within the sample. It has been observed that the proportional analysis between FTIR spectra and histopathological investigations of normal and tumor breast cells showed that FTIR spectroscopy is a trustworthy method for tumor diagnosis [28].

Circular Dichroism (CD)

Proteins that have been purified from tissues or obtained using recombinant techniques are studied by Circular Dichroism (CD) which is a remarkable instrument for rapid determination of the secondary structure and folding characteristics of proteins. It is well established that rapid categorization of new proteins is of great significance for the fields of proteomics and structural genomics. With the help of CD multiple samples containing 20μg or less of proteins in physiological buffers could be measured in a few hours. On the other hand, it does not give the residue specific data/information that can be obtained by NMR or X-ray crystallography. Principally, it works on the unequal absorption of left-handed and right-handed circularly polarized light.

Raman spectroscopy

The structure of unfolded polypeptides is studied by Raman spectroscopy. Raman spectroscopy has the benefit of several essential

advantages in characterizing the vibrational spectra and secondary structural affinity of native unfolded proteins. Raman spectra can be achieved in dilute aqueous solution, a good feature that is of enhanced significance in characterizing natively unfolded proteins because of their inclination to aggregate at higher concentrations. Moreover, Raman spectra permit the environment of several amino acid side chains to be characterized, including the acidic residues, sulfur-containing residues and aromatic amino acids in different physical states.

Dual Polarization Interferometry (DPI)

Dual polarization interferometery is greatly sensitive analytical methodology employed for the determination of structure, function as well as orientation of biological and other layers at solid-liquid interface, providing the measures of multiple parameters of molecules at a surface showing information associating with packing and thickness related to layer refractive index (RI), stoichiomety (mass), surface loading and density. DPI has been employed to investigate multiple types of phospholipids. Moreover, peptides relating to plasma membrane like duramycin and V4 (antimicrobial peptide) and Melitti (component of bee venom) have been scrutinized (interaction between meittin-lipid) by DPI. Mass and structural changes, in various proteins due to the function of metal ion concentration, including Ca^{2+} in calmodulin, prion containing Fe^{2+}, β-amyloid associated with Cu^{2+} etc have been studied by DPI [29]. DPI methodology has also been employed for evaluating immobilization approach for DNA sensing surface and is challenging because of high charge on DNA backbone rendering reachable orientation of molecule hard to accomplish [30,31].

Crystallographic analysis

As mentioned in the section of "Introduction", x-ray crystallography is a type of high resolution microscopy which facilitates the researchers to scrutinize the protein structures at the protein level and enhances the insight into protein functionality. As far function is concerned, interaction of protein with other molecules, process of catalysis in case of enzymes and changes in conformational alteration and studied during x-ray crystallographic analysis.

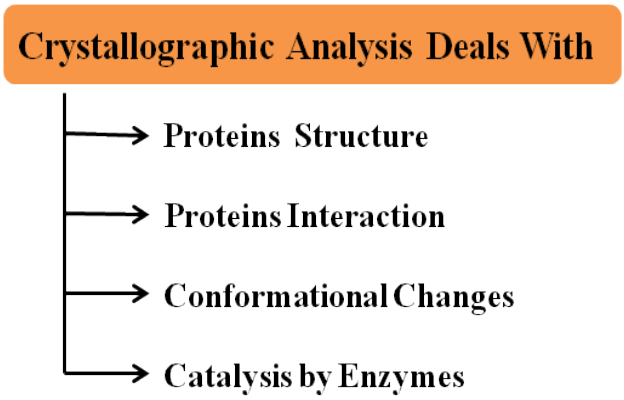


Figure 11: Use of Crystallography in Protein Structure and Function Determination.

In all types of microscopy, the detail and/or resolution depends upon the wavelength (λ) of the electromagnetic radiation used. Light microscopy (λ= ~300nm) shows individual cells and its components (sub-cellular organelles). On contrary, electron microscopy (λ= <10nm) helps to visualize cellular structures and large protein molecules easily. While to observe the protein molecules at atomic level, x-rays are employed having wavelength about 0.1nm or 1Å. In light microscopy, the incident radiation is diffracted in all directions, the diffracted beams are magnified by the lenses and the enlarged image is obtained of the specimen. On contrary, in electron microscopy, the diffracted beams are collected by magnets. While in x-ray analysis, the diffracted pattern cannot be focused physically. The patterns produced during x-ray analysis is recorded by the detector using x-ray sensitive film, usually a charge-coupled device is also used.

The diffraction pattern is too weak to be measured. Hence, an ordered three-dimensional (3D) array (crystal) of molecules is employed to magnify the signal. Therefore, magnification or detailed structure information in the form of high resolution depends upon the internal order of molecules in the protein crystal, even a small protein or polypeptide molecules contain thousands of atoms in it. So, orderly arranged molecules in crystal will result in high resolution structure with crystallographic analysis and vice versa. The x-ray diffraction pattern forming 3D map illustrates distribution of electrons in the structure (crystal). The effect of constructive and destructive interference is produced by crystal because it behaves like a 3D diffraction grating and on detector as series of discrete spots are observed, known as reflections. Every reflection gives the information of all atoms in the structure. Being an electromagnetic in nature, x-rays behave wave like characteristics having both the amplitude and a phase.

The phase problem: The amplitude is recoded easily from diffraction pattern but information of phase is lost, and known as the phase problem. When the structure of protein is resolved, it is said that phase problem has resolved. In other words, phase data/information is satisfactory to illustrate electron density map.

Protein preparation: A pure sample of target protein is the first requirement. It could be obtained by separation from its source or by gene into high expression system. The following steps are applied on sample to check its purity:

- Purity and homogeneous test is applied by electrophoresis and mass spectrometry methods.
- Protein solubility and foldness is examined by circular dichroism (CD).
- Monodispersion of sample: aggregation in the sample must be avoided by employing dynamic light scattering (DLS) device.
- Activity assay is applied to check the activeness of protein.
- Sample stability: protein crystals are formed at room temperature, overnight, but usually it takes many days.

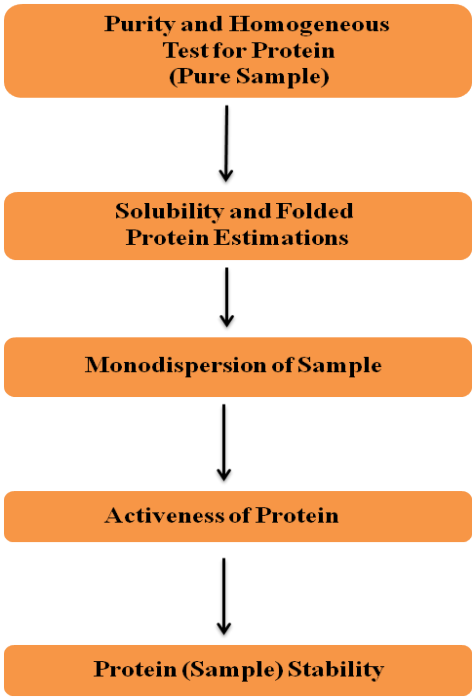


Figure 12: Steps in Protein Preparation and Analysis.

X-Ray data collection: In x-ray crystallographic analysis, several aspects are studied for a given crystal of sample (protein). The resolution limit, the unit cell parameter, crystal symmetry and crystal orientation are studied usually. With the help of these parameters' information, a data collection approach is formed. During collection of data set of typical medium resolution, may take up to three days by employing x-ray source. While for high resolution data, a synchrotron radiation is utilized, where the intensity of x-rays is greater and time is shorter in case of data collection.

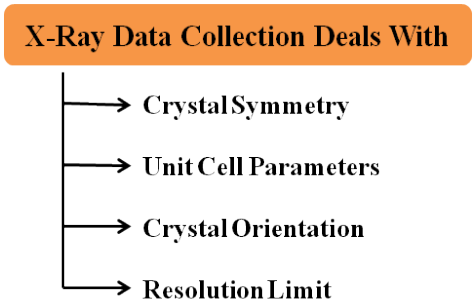


Figure 13: Use of X-Ray Data in Determination of Various Aspects of a Crystal.

Electron crystallography: Results have shown that electron crystallography of two dimensional crystals has progressed to atomic resolution. Verities of proteins have been crystallized in two dimensional format and conformation [32]. The structure obtained by electron crystallography like tubulin, the light harvesting complex -2 (LH2), aquaporin-1 (AQP1) and bacteriorhodopsin (BR) have allowed significant atomic models to be built [33-38].

Many membranous proteins have been examined by electron crystallography and their structures studied to a resolution that produced the secondary structure to be clearly illustrated. This technique has advantage over the others because it provides fast data collection and processing.

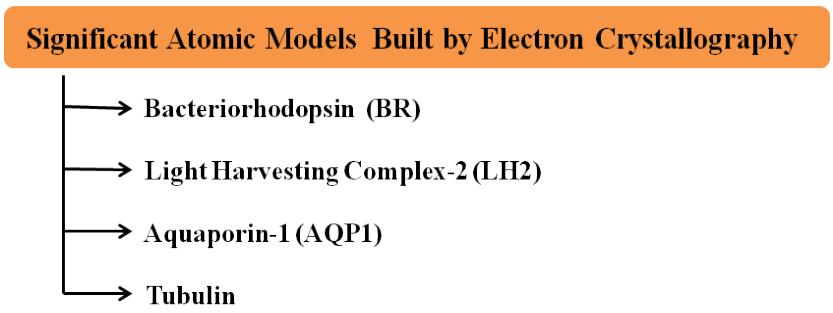


Figure 14: Significance of Electron Crystallography in the Formation of Atomic Models of Biological Molecules.

Atomic Force Microscopy (AFM): Atomic force microscopy (AFM) allows the investigation of two-dimensional (2D) membrane

protein crystals insight into the structure and dynamics at micro level resolution like electron crystallography. With the help of AFM detailed information of structure and dynamics of membrane protein could be assessed.

Cryoelectron microscopy: Cryoelectron microscopy produces 3D structures of proteins at atomic resolution. The combined application of cryoelectron microscopy with other microscopic techniques established 3D structure of different membrane proteins and allowed the visualization of conformational changes during catalytic cycles.

Computational Methods: Molecular Simulation Techniques

Computational based methods present an improvement for perceptive protein flexibility and stability as well. However, Molecular Dynamics (MD) simulation models render interactions among atoms of polypeptide chain.

The Folding and unfolding pathways of Proteins
Formation of the final native structure
Time dependence of pathways (folding and unfolding)
The inter-residue interactions that underlie the processes

Table 1: Advantage of Molecular Dynamics (MD).

Below is the list of molecular simulation techniques (Table 02) that are employed to investigate the various aspects of protein dynamics especially folding and unfolding, *in silico*.

S.#	Tools	Particulars
1	ModView	For the illustrating multiple protein sequences and structural analysis
2	Perturbation Analyzer	Tool study for the protein network interactions
3	Protein Analyst	Distributed object environment for protein sequence and structure analysis
4	VISTRAJ	Exploring protein conformational space
5	PROTMAP2D	Visualization, comparison and analysis of 2D maps of protein structure
6	FlexServ	Integrated tool for the analysis of protein flexibility
7	iFold	Platform for interactive folding simulations of proteins
8	Bio3d	An R package for the comparative analysis of protein structures
9	PConPy	Python module for generating 2D protein maps
10	COREX/BEST server	Regional stability and thermodynamic aspects
11	CAVER 3.0	For the study of transport pathways
12	ExPASy (Expert Protein Analysis System)	Bioinformatics resource portal
13	UniProt	Database of protein sequence and functional information
14	RasMol	For molecular graphics visualization
15	BLAST	For comparing primary biological sequence information
16	T-coffee	Multiple sequence alignment software

Table 2: Molecular Simulation Techniques.

ModView

For the visualization of multiple protein sequences and structure analysis, a program has been introduced called ModView (a web based application). It facilitates the users with data base query engine, multiple sequence alignment editors, and multiple structure viewers. Users can control hundreds of proteins interactively, to create fragments, active and binding sites as well as domains in protein families to show macromolecular complexes of virus or ribosome. ModView, has been incorporated as plug-in in Netscape (an internet browser), along with text and figures that could be used for teaching and presentations [39].

Perturbation analyzer

Perturbation analyzer has been developed for using with Cytoscape (a program for visualizing molecular interaction networks and integrating these interactions with gene expression profiles). Basically it is an open source plug-in and could be used in manual approach for simulating user defined perturbations, and help for estimating the network robustness and identification of significant proteins that produce large effects in protein network interactions (PINs), when concentrations are perturbed. It has a great deal in exploring the design principles of protein networks and could be a valuable for identifying drug target [40].

Protein Analyst

Protein Analyst is a tool for the analysis of protein sequences with prominence on the integration of sequence and structural information. Protein Analyst is considered to be a flexible integrated tool for delivering existing and newly developed protein bioinformatics algorithms to the desktop with emphasis on the integration of sequence and structural information [41].

VISTRAJ

VISTRAJ allows 3D visualization, handling and editing of protein conformational space with the help of probabilistic maps of the space. These probabilistic maps known as trajectories distributions and serve as input to FOLDTRAJ, while FOLDTRAJ generates sample protein structures based on conformational space. Both could be used as tools for homology model formation and 3D structures are produced containing post translational modified amino acids. FOLDTRAJ helps in generating plausible probabilistic protein conformers [42].

PROTMAP 2D

Protein structure comparison is an elementary issue in the field of structural biology and bioinformatics. 2D maps regarding distances between residues contain information to restore the 3D representation. The maps disclose patterns of interaction between secondary and super-secondary structures and are helpful in visual investigations. The overlapping of the maps of to structures gives a measurement of protein structure similarity. PROTMAP 2D helps in calculation of contact and distance maps, based on user defined parameters, quantitative measurement of pairs or series of contact maps and visualization of the results [43].

FlexServ

The flexibility of proteins can be analyzed by FlexServ (a web-based program). The incorporated protocols in the server like normal mode analysis (NMA), discrete molecular dynamics (DMD) and Brownian dynamics are used for protein dynamics. User defined trajectories could also be analyzed. The server facilitates all the parameters related to the flexibility analysis e.g. temperature factors, basic geometrical analysis, inflexibility analysis, chain correlation, dynamic domain determination, essential dynamics etc. The complete data is illustrated by plain text, 2D and 3D graphics [44].

iFold

Molecular simulations of protein dynamics could be performed by using iFold (web-based program). This server provides various features regarding protein folding of large scale simulations, thermodynamic analysis, pfold analysis and simulated annealing with the help of discrete molecular dynamics (DMD). Moreover, protein models could be generated by interactions among amino acids [45].

Bio3d

The homologous structure of proteins is analyzed by using a computer program, Bio3d. It has great advantage of graphics as well as statistics. Homologous analysis helps in the study of internal conformational differences of structures and inter-conformer associations. Moreover, it facilitates the exploring of structural protein evolution. Various features of this program assist in exploring structure and sequences of proteins like dynamic trajectory data, sequence data, re-orientation, atom selection, sequence conservation analysis, distance matrix analysis, clustering analysis, principal component analysis (PCA) and rigid core analysis [46].

PConPy

An open-source Python (A simple, high-level interpreted language) module for creating protein contact maps, distance maps and hydrogen bond plots. Contact maps can be explained with secondary structure and hydrogen bond assignments. These maps can be produced in a number of publication-quality vector and raster image formats. PConPy offers a more flexible choice of contact description parameters than existing toolkits. PConPy can be employed as stand-alone software or imported into existing source code (source language). A web-interface to PConPy is also available for use [47].

DYNATRAJ

DYNATRAJ facilitates visualization of modes of motion and PCA for ensemble of protein structures taken from the molecular dynamics and experimental ensembles by using NMR.

DYNAPOCKET: DYNA POCKET provides configuration of a protein pocket found in single protein structure. It helps in prediction of drug designing by using binding pocket conformations.

COREX/BEST Server

This server has ability to generate structural and thermodynamic aspects of protein structures. The conformational ensemble provides stability of protein structure and conformation. The stability of protein conformations are expressed by kcal/mol (units of energy). The stability of regions in the structure at the resolution of residues is mapped onto protein structure for visual expression [48].

CAVER

CAVER is extensively employed for the study of transport pathways in protein macromolecular structures. The idea of tunnel and channels is crucial for the understanding of transport of small particles (polar or non-polar) through plasma membrane. The perceptiveness of association between structure and functionality of proteins helps in designing of new inhibitors and production of biocatalysts. CAVER 3.0 helps in the analysis of tunnel and channels in protein conformations. It also gives insight into clustering in pathways [49].

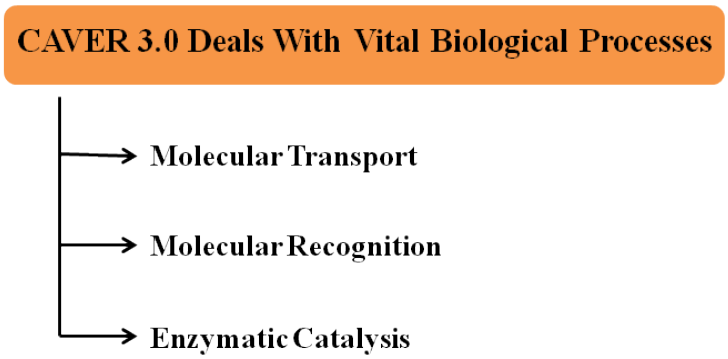


Figure 15: Caver 3.0 Paves the Technique for the Study of Essential Biochemical Mechanisms.

ExPASy (Expert Protein Analysis System)

ExPASy is a bioinformatics resource portal (gateway) operated by the Swiss Institute of Bioinformatics (SIB). It is an integrative portal accessing many databases, software tools and scientific resources in different areas of life sciences (proteomics, phylogeny/evolution, genomics, transcriptomics, systems biology, population genetics, etc). ExPASy acts as a proteomics server to analyze protein sequences and structures and two-dimensional (2D) gel electrophoresis [50].

UniProt

UniProt is a freely accessible database of protein functional information and sequence as well as many entries being derived from genome sequencing projects. It consists of a large amount of information about the biological function of proteins derived from the

research literature. UniProt provides four core databases: UniProtKB (with sub-parts Swiss-Prot and TrEMBL), UniParc, UniRef, and UniMes [51].

RasMol

RasMol is used for molecular graphics visualization intended and primarily for the depiction and exploration of biological macromolecule structures. RasMol has an important educational tool as well as ongoing to be an important tool for research in structural biology.

BLAST (Basic Local Alignment Search Tool)

It is an algorithm for comparing primary biological sequence information, such as the nucleotides or amino-acid sequences. BLAST search enables a researcher to compare a query sequence with database of sequences, and recognize sequences that are similar to the query sequence [52]. BLAST can be used for several purposes.

Purpose	Particulars
Identifying Species	For the identification of species and homologous sequence
Locating Domains	While working with sequences, identification of domains
Establishing Phylogeny	Create a phylogenetic tree
DNA Mapping	Looking for gene sequence at unknown place in known species
Comparison	Locating common sequences in species while working on genes

Table 3: Applications of Blast.

T-Coffee (Evaluation of alignment)

Multiple sequence alignment is studied by T-Coffee. Sequences could be merged into each other and library of pair wise alignments could be generated. Structural data could be obtained from protein data bank regarding sequence alignment [53]. Various modes of T-coffee are listed below:

Mode	Particulars
M-Coffee	It combines the results of different multiple sequence analyzing programs
Expresso and 3D-Coffee	It combines sequences and structures in particular alignment
R-Coffee	RNA sequences could be aligned
PSI-Coffee	Proteins are aligned by homology extension
TM-Coffee	Transmembrane proteins are aligned
Pro-Coffee	Promoter regions are aligned

Table 4: Various Modes of T-Coffee.

References

1. Richards FM (1977) Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng 6: 151-176.
2. Montell C, Devreotes P (2004) From protein dynamics to animal behavior: new insights into complex cell regulatory mechanisms. Curr Opin Cell Biol 16: 115-118.
3. Weiss M, Hashimoto H, Nilsson T (2003) Anomalous protein diffusion in living cells as seen by fluorescence correlation spectroscopy. Biophys J 84: 4043-4052.
4. Bacia K, Scherfeld D, Kahya N, Schwille P (2004) Fluorescence correlation spectroscopy relates rafts in model and native membranes. Biophys J 87: 1034-1043.
5. Douglass AD, Vale RD (2005) Single-molecule microscopy reveals plasma membrane microdomains created by protein-protein networks that exclude or trap signaling molecules in T cells. Cell 121: 937-950.
6. Kusumi A, Nakada C, Ritchie K, Murase K, Suzuki K, et al. (2005) Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule tracking of membrane molecules. Annu Rev Biophys Biomol Struct 34: 351-378.
7. Pike LJ (2006) Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function. J Lipid Res 47: 1597-1598.
8. Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. Science 175: 720-731.
9. Gohlke H, Kuhn LA, Case DA (2004) Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. Proteins 56: 322-337.
10. Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, et al. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305: 386-389.
11. Rasmussen BF, Stock AM, Ringe D, Petsko GA (1992) Crystalline ribonuclease A loses function below the dynamical transition at 220 K. Nature 357: 423-424.
12. Cui Q, Li G, Ma J, Karplus M (2004) A normal mode analysis of structural plasticity in the biomolecular motor F(1)-ATPase. J Mol Biol 340: 345-372.
13. Vitagliano L, Merlino A, Zagari A, Mazzarella L (2002) Reversible substrate-induced domain motions in ribonuclease A. Proteins 46: 97-104.
14. Fayer MD (2001) Fast protein dynamics probed with infrared vibrational echo experiments. Annu Rev Phys Chem 52: 315-356.
15. Schlessinger J (1988) Signal transduction by allosteric receptor oligomerization. Trends Biochem Sci 13: 443-447.
16. Yap AS, Brieher WM, Gumbiner BM (1997) Molecular and functional analysis of cadherin-based adherens junctions. Annu Rev Cell Dev Biol 13: 119-146.
17. Perez-Moreno M, Jamora C, Fuchs E (2003) Sticky business: orchestrating cellular signals at adherens junctions. Cell 112: 535-548.
18. Pujuguet P, Del Maestro L, Gautreau A, Louvard D, Arpin M (2003) Ezrin regulates E-cadherin-dependent adherens junction assembly through Rac1 activation. Mol Biol Cell 14: 2181-2191.
19. Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. Chem Rev 106: 3055-3079.
20. Benkovic SJ, Hammes-Schiffer S (2003) A perspective on enzyme catalysis. Science 301: 1196-1202.

21. van der Kamp MW, Shaw KE, Woods CJ, Mulholland AJ (2008) Biomolecular simulation and modelling: status, progress and prospects. J R Soc Interface 5 Suppl 3: S173-190.

22. Beck DA, Daggett V (2004) Methods for molecular dynamics simulations of protein folding/unfolding in solution. Methods 34: 112-120.

23. Cui M, Shen J, Briggs JM, Fu W, Wu J, et al. (2002) Brownian dynamics simulations of the recognition of the scorpion toxin P05 with the small-conductance calcium-activated potassium channels. J Mol Biol 318: 417-428.

24. Yu L, Sun C, Song D, Shen J, Xu N, et al. (2005) Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex. Biochemistry 44: 15834-15841.

25. Meng XY, Xu Y, Zhang HX, Mezei M, Cui M (2012) Predicting protein interactions by Brownian dynamics simulations. J Biomed Biotechnol 2012: 121034.

26. Angarica VE, Sancho J (2012) Protein dynamics governed by interfaces of high polarity and low packing density. PLoS One 7: e48212.

27. Kong J, Yu S (2007) Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim Biophys Sin (Shanghai) 39: 549-559.

28. Dimitrova M, D Ivanova, I Karamancheva, A Milev, I Dobrev (2009) Application of FTIR-Spectroscopy for Diagnosis of Breast Cancer Tumors. Journal of the University of Chemical Technology and Metallurgy 44: 297-300.

29. Karim K, JD Taylor, DC Cullen, MJ Swann, NJ Freeman (2007) Measurement of conformational changes in the structure of transglutaminase on binding calcium ions using optical evanescent dual polarisation interferometry. Analytical Chemistry 79: 3023-3031.

30. Berney H, Oliver K (2005) Dual polarization interferometry size and density characterisation of DNA immobilisation and hybridisation. Biosens Bioelectron 21: 618-626.

31. Lillis B, Manning M, Berney H, Hurley E, Mathewson A, et al. (2006) Dual polarisation interferometry characterisation of DNA immobilisation and hybridisation detection on a silanised support. Biosens Bioelectron 21: 1459-1467.

32. Ringler P, Heymann BJ, Engel A (2000) Two-dimensional crystallization of membrane proteins. Edited by Baldwin SA: Oxford University Press, Oxford.

33. Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, et al. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. J Mol Biol 213: 899-929.

34. Mitsuoka K, Hirai T, Murata K, Miyazawa A, Kidera A, et al. (1999) The structure of bacteriorhodopsin at 3.0 A resolution based on electron crystallography: implication of the charge distribution. J Mol Biol 286: 861-882.

35. Kühlbrandt W, Wang DN, Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. Nature 367: 614-621.

36. Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, et al. (2000) Structural determinants of water permeation through aquaporin-1. Nature 407: 599-605.

37. Ren G, Cheng A, Reddy V, Melnyk P, Mitra AK (2000) Three-dimensional fold of the human AQP1 water channel determined at 4 A resolution by electron crystallography of two-dimensional crystals embedded in ice. J Mol Biol 301: 369-387.

38. Nogales E, Wolf SG, Downing KH (1998) Structure of the alpha beta tubulin dimer by electron crystallography. Nature 391: 199-203.

39. Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, McMahan L, et al. (2003) ModView, visualization of multiple protein sequences and structures. Bioinformatics 19: 165-166.

40. Li F, Li P, Xu W, Peng Y, Bo X, et al. (2010) Perturbation Analyzer: a tool for investigating the effects of concentration perturbation on protein interaction networks. Bioinformatics 26: 275-277.

41. Saqi MA, Wild DL, Hartshorn MJ (1999) Protein analyst--a distributed object environment for protein sequence and structure analysis. Bioinformatics 15: 521-522.

42. Salama JJ, Feldman HJ, Hogue CW (2001) VISTRAJ: exploring protein conformational space. Bioinformatics 17: 851-852.

43. Pietal MJ, Tuszynska I, Bujnicki JM (2007) PROTMAP2D: visualization, comparison and analysis of 2D maps of protein structure. Bioinformatics 23: 1429-1430.

44. Camps J, Carrillo O, Emperador A, Orellana L, Hospital A, et al. (2009) FlexServ: an integrated tool for the analysis of protein flexibility. Bioinformatics 25: 1709-1710.

45. Sharma S, Ding F, Nie H, Watson D, Unnithan A, et al. (2006) iFold: a platform for interactive folding simulations of proteins. Bioinformatics 22: 2693-2694.

46. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS (2006) Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22: 2695-2696.

47. Ho HK, Kuiper MJ, Kotagiri R (2008) PConPy--a Python module for generating 2D protein maps. Bioinformatics 24: 2934-2935.

48. Vertrees J, Phillip Barritt, Steve Whitten, Vincent J Hilser (2005) COREX/BEST server: a web browser-based program that calculates regional stability variations within protein structures. Bio-Informatics 21: 3318–3319.

49. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, et al. (2012) CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput Biol 8: e1002708.

50. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, et al. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31: 3784-3788.

51. UniProt Consortium (2011) Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res 39: D214-219.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

53. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302: 205-217.

Advances in Protein Chemistry

Chapter: Advances in Protein Thermodynamics

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published September, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Advances in Protein Thermodynamics

Arif Malik[1], Mahmood Rasool[2*], Abdul Manan[1], Aamer Qazi[3] and Mahmood Husain Qazi[4]

[1]Institute of Molecular Biology and Biotechnology (IMBB), The University of Lahore, Pakistan

[2]Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

[3]Ontario Institute for Cancer Research, Toronto, Canada

[4]Center for Research in Molecular Medicine (CRiMM), The University of Lahore-Pakistan

***Corresponding author:** Dr. Mahmood Rasool, PhD Assistant Professor, Center of Excellence in Genomic Medicine Research, Post Box No. 80216, King Abdulaziz University, Jeddah 21589, Saudi Arabia, Tel: +966-582-254267; E-mail: mahmoodrasool@yahoo.com

Introduction

Proteins are synthesized in the cytoplasm or *in vitro* as amorphous polypeptide chains that, usually, assemble into their functionally active three-dimensional (3D) shapes, a process; known as protein folding. In physico-chemical perspective, it is achievable in characterizing the folding mechanism of a given protein at molecular as well as atomic level and to restructure its free-energy landscape. Biological systems abide by the natural laws of chemistry and physics. The gradual progress in the field of biological sciences involve the information on various mechanisms/action that have affected the living systems at the molecular level during the investigations of the biological mechanisms in turn the structure of several molecules like proteins and nucleic acids have been determined. The physical sciences, including thermodynamics, can make a crucial contribution to the biological sciences for understanding of various biological mechanisms.

Thermodynamics is distinct by the energy level distribution; the dilemma lies in the production of a pragmatic/practical protein-like model that would suitably take hold of most of thermodynamic properties of proteins like the denaturation temperature, **Entropy**, **enthalpy**, heat capacity. Thermodynamics plays an important part in driving the dynamics of protein folding. It is well recognized that the greater part of proteins can achieve their native states and they can also refold to their native states after been denatured. To account for these astonishing properties the familiarity of the energy level distribution is not satisfactory, in other words the distribution of energy in forming the new bonds between the residues of protein need to be explored at molecular as well as atomic levels. It turns out that the dynamics of protein folding is much more complicated to understand regarding protein thermodynamics. Simple systems quickly develop towards equilibrium, because an isolated system is distinguished by utmost **Entropy** (Disorder). On the contrary, the living system does not reach at the state of stability or equilibrium easily. The biological development shows the increasing trend of growth of multi-cellular organism from the unit cell. The thermodynamic studies of such a complicated system in various organisms are really a big challenge to biologists.

Equilibrium thermodynamics helps for the study of biological processes *in vitro* environment. Equilibrium characteristics of biological macromolecules are one of the basic requirements and essential for the researchers to know about them. Moreover, the combination of statistical and molecular models with equilibrium thermodynamics provides microscopic understanding of various mechanisms operating the living systems. In receipt of above mentioned mechanisms there are some processes regarding binding between biological macromolecules in cells or between a macromolecule and a ligand outside the cell. Various complex events take place in the biological system; they comprise active diffusion, translocation of proteins, conformational changes, membrane fusion, virus assembly, vesicle budding and DNA unwinding. Moreover, posttranslational modification (PTM), a crucial process, is everywhere in the cell and control the function of proteins frequently by modulating their biophysical properties. These modifications are methylation, glycosylation, acetylation, ubiquitination, S-nitrosylation, lipidation and phosphorylation which can regulate the structural features of proteins thermodynamic and kinetic.

The current chapter has main two parts one based on practical/experimental approach and the other on computational approach in which tools of bioinformatics and various databases have been mentioned which are used to study various aspects of proteins including their thermodynamic aspects. Thermodynamics deals with the association between heat and work. Proteins, polymer of amino acids, exhibit several crucial physiological behaviors and have unique structural characters like stability. Environmental conditions are also crucial factors in the folding of proteins, since a change in conditions can annihilate the native (Folded) structure. The folding and unfolding mechanisms are described in terms of thermodynamics.

S r.no	Various Applications of Thermodynamics in Proteins
1	Stability of proteins
2	Improvement of bioprocess
3	Biomass and metabolite production
4	Cell transport
5	Equilibrium studies in downstream process
6	Properties of biomolecules

Study of thermodynamics depends upon the basic physical laws of thermodynamics. The first law (law of energy conservation) deals

with the total energy of an isolated system are constant despite internal changes. Moreover, second law which is related to **Entropy** of the system, states the mechanical work results when a body interacts with another body at lower temperature. Therefore, any autonomous process shows an increase in **Entropy**. Another important law that deals with thermal equilibrium (zeroth law of thermodynamics) states that if two objects are in thermal equilibrium with a third object then the first two objects would be in thermal equilibrium with respect to each other. The Gibbs free energy (**ΔG**) equation is helpful in understanding of thermodynamic activities of a system. In general, free energy is a thermodynamic quantity equivalent to the capacity of a physical system to do work. Thermodynamics of protein stability reveals a general tendency for proteins that denature at higher temperatures to have greater free energies of maximal stability. There is a constant equilibrium between proteins in denatured and native states. The conditions in the system determine the amount of one of the species. There are usually one or few states of folded proteins and numerous states of unfolded proteins.

$$\Delta G_F < \Delta G_D$$

All the systems desire for the lowest energy possible. There is a barrier between denatured and folded protein, a transition state, which needs to be overcome to shift from one form to the other. Mathematical relation of thermodynamic parameters in the form of Gibbs free energy equation is below:

$$\Delta G = \Delta H - T\Delta S$$

Where, **ΔG-free energy** is measure of protein stability. The more negative is the free energy, the more stable structure of protein is observed, **ΔH– enthalpy** is bond formation and breaking in protein while **ΔS-Entropy** is degrees of freedom in protein. Free Energy (**ΔG**) is a thermodynamic quantity equivalent to the capacity of a physical system to do work. **Entropy** (**ΔS**), quantity representing the amount of energy in a system that is no longer available for doing mechanical work, favors unfolding of protein, and you have to overcome its barrier by lowering the **ΔG**. On the other hand, **Enthalpy** (**ΔH**), a thermodynamic quantity equal to the internal energy of a system plus the product of its volume and pressure, favors folding of protein, because the number of inter-atomic interactions increases upon protein folding. **Enthalpy** (**ΔH**) and **Entropy** (**ΔS**) are about the same size usually, so they cancel each other showing proteins are not particularly stable, which is very imperative for their biological function. The resulting **ΔG** is therefore rather small, although **ΔH** and **ΔS** themselves are large.

Forces that Govern Protein Stability

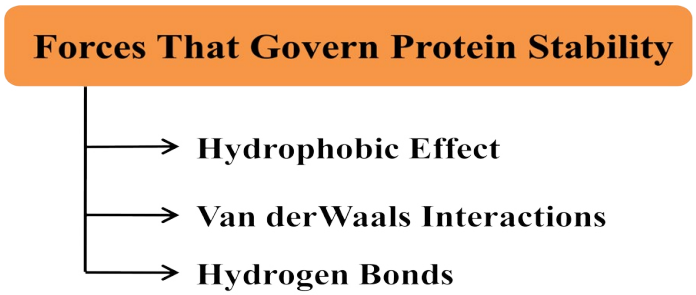


Figure 1: Forces in Atoms of Amino Acids that governs Protein Stability.

Hydrophobic effect: This is the most significant force in protein stability. Hydrophobic residues tend to associate and turn to the inside of the core of the protein molecule. Hydrophilic residues will face the outside (in contact with water). For example, water molecules surround the oil and lose some degrees of freedom. When hydrophobic residues (oil) are buried inside the protein, the disorder of water increases, and **ΔG** decreases- the reaction is more spontaneous.

Van der Waals (VDW) interactions: these are relatively weak attractive forces between neutral atoms and molecules arising from polarization induced in each particle by the presence of other particles. Inside the protein, the residues are tightly packed due to Van der Waals interactions. The types of Van der Waals interactions govern the stability of proteins e.g. London dispersion forces and dipole-dipole forces.

Van der Waals potential: If the atoms are too far, the interaction between them is not observable. If they are too close- they collide with each other. It should be noted that there is some optimum distance in which attractive and repulsive forces are in equilibrium.

Hydrogen bonds: Hydrogen bonds are interaction between a hydrogen atom attached to an electronegative atom and a lone pair of electrons (Also on an electronegative atom).

Practical/Experimental Approach

Fluorescence: The amino acid Tryptophan is hydrophobic in nature, if it is buried (inside the folded or native protein) the fluorescence is small. In principle, when it is exposed to the surface (for instance in the unfolded state) fluorescence is increased, as a result.

Calorimetery: Heat the protein and measure denaturation over different temperature ranges. When protein unfolds, it produces heat. Hence, calorimeter measures the released heat gives the indication for denaturation.

CD Spectroscopy: There are characteristic absorption spectra for different protein structures. Proteins of approximately 300 residues other than small globular structures have been most studied and the results cannot be applicable to other larger proteins like fibrous proteins and/or those present in plasma membrane. However, the basic rules of protein folding determined during the study of similar globular proteins, could be applicable on the more complex protein structure and mechanism of folding could be revealed properly.

Other various techniques for analyzing protein dynamics:

1. Fluorescence recovery after photobleaching (FRAP)

2. Single particle tracking (SPT)

3. Fluorescence correlation spectroscopy (FCS)

4. Nuclear Magnetic Resonance (NMR)

5. X-Ray and Neutron Scattering

6. Fluorescence Technique

7. Single Molecule Technique

8. Hydrogen Exchange Mass Spectrometry

9. Fourier Transform Infrared (FTIR) Spectroscopy

10. Circular Dichroism (CD)

11. Raman Spectroscopy

12. Dual Polarization Interferometry (DPI)

13. Crystallographic Analysis

• Electron Crystallography

• Atomic Force Microscopy

• Cryoelectron Microscopy

Brief description is in the following paragraphs:

• Imaging methods such as fluorescence recovery after photo-bleaching (FRAP), fluorescence resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCS) have been modified so that they can be done on commercially available and user-friendly laser scanning microscopes.

• Nuclear magnetic resonance (NMR) spectroscopy is employed to scrutinize the dynamic behavior of a protein at a multitude of specific sites. Moreover, protein movements on a broad range of timescales can be screened using various types of NMR experiments. The high resolution NMR spectroscopy is found to obtain comprehensive information about the structure and dynamics of proteins, in order to explicate their functions.

• Neutron scattering and X-ray analysis provide elaborative description on protein dynamics.

• Fluorescence provides understanding of protein dynamics both in single molecule and in ensemble form. The experiments of single molecular study are informative, and help in the understanding of dynamics of individual molecules and give insight how translation occurs into an ensemble signal.

• Single molecule technique presents the exciting likelihood of observing individual protein dynamics within a true cellular framework.

• Hydrogen exchange together with mass spectrometry (MS) is a precious analytical tool for the study of protein dynamics. Information about protein dynamics, by combining, with more classical functional data, a more systematic understanding of protein function can be obtained.

• Infrared (IR) spectroscopy is one of the well recognized experimental techniques for the investigation of secondary structure of polypeptides as well as proteins. IR spectrum can be achieved for proteins in a large range of environments with a small amount of sample. IR gives information on protein dynamics and structural stability.

• Proteins that have been purified from tissues or obtained using recombinant techniques are studied by Circular Dichroism (CD) which is a remarkable instrument for rapid determination of the secondary structure and folding characteristics of proteins.

• The structure of unfolded polypeptides is studied by Raman spectroscopy. Raman spectroscopy has the benefit of several essential advantages in characterizing the vibrational spectra and secondary structural affinity of native unfolded proteins.

• Dual polarization interferometry (DPI) is greatly sensitive surface analytical technique that has been employed for measuring the functionality, structure and orientation of biological and other layers at the liquid-solid interface, providing the measurement of various parameters of molecules at a surface showing information on molecular dimension associating with layer thickness and packing related to layer refractive index (RI) and density as well as surface loading and stoichiometry (mass).

• X-ray crystallography is essentially a form of very high resolution microscopy. It facilitates in research to visualize protein structures at the atomic level and enhances our understanding of protein function.

• Electron crystallography allows the study of two-dimensional membrane protein crystals. X-ray crystallography can give good structural information by allowing the determination of the average position of atoms and the amplitudes of their displacements from the average positions.

• Atomic force microscopy, like electron crystallography, allows the study of two-dimensional membrane protein crystals. Atomic force microscopy gives insight into the surface structure and dynamics at sub-nanometer resolution.

• Cryoelectron microscopy allows the 3D structure of the vitrified protein to be assessed at atomic resolution.

Folding process of protein in a biological system depends upon the various factors that govern the folding into complex form/structure like polarity of protein molecules, hydrophobic and hydrophilic effects and association between protein molecules and surrounding solvent. In aqueous environment, polar amino acids exhibit hydrophilic nature, exert a pull on polar water molecules while non-polar amino acids are inclined to be hydrophobic. The hydrophobic portion of protein residues do not show any interaction with water rather associates with other residues [1]. The folded conformation of protein is also associated with peptide linkages present between the consecutive amino acids. The folding process may covers the alignment of intermediate structures that could be larger than the native one and showing integral secondary structures, these all termed as molten globule [1-2].

The finishing stage of folding depends on the precise and specific sequence of amino acids, while previously folding stages are supposed to be mostly insensitive to information of sequence [3].

Thermodynamics of Protein Conformational Changes

The folded conformation seems to be at narrow free energy (**ΔG**) state but any change in the folded conformation depends upon the significant increment in free energy. The change in heat capacity when protein unfolding occurs is the temperature at which the stability is high in its value. As far as free energy is concerned, the highest stability occurs when S=0, on the other hand calculation by the equilibrium constant appears when **enthalpy** change is zero. The maximum stabilities could be occurred at different temperature ranges, but both are considered in different cases. The stabilities of native form decrease at both lower and higher temperatures.

The other factors such as binding interactions among the amino acids seem to be playing an elaborative role in the stability of the protein conformation but these are not responsible for the significant stability and similar situation happens in the unfolded state, while the bonding between native protein and solvent is expected to be strong enough as compared to the bonding between unfolded coil and the surrounding environment. Moreover, the hydrophobic effect is considered to be the major stabilizing player.

Energy Landscape Theory and Thermodynamics of Protein Stability

Protein folding is one of the complex processes. Frauenfelder et al. and Bryngelson et al. describe the complexity of protein folding by restoring to a statistical approach to the energetic of protein conformation in the form of energy landscape [4-5]. Bryngelson et al. refers it mathematical procedure that assist in understanding of microscopic behavior of molecular system. It gives quantitative explanations of folded protein state, ensembles of conformational sub-states, ensembles of folding intermediates and denatured or unfolded states and considered as the realistic model of protein [6-7]. Bryngelson et al. describes an energy landscape, in mathematical form, of a system with "n" degrees of freedom is an energy function [5]:

$$F(x) = F(x_1, x_2, x_3, ..., x_n)$$

Here, $x_1, x_2, x_3, \ldots, x_n$ are variables for the microscopic state of the system [8] and F(x) is defined as the free energy. As far as protein is concerned, these variables ($x_1, x_2, x_3, \ldots, x_n$) are all the dihedral angles of the chain and showing a single conformation of protein. The stability of the protein can be determined by examining the set of values $x_1, x_2, x_3, \ldots, x_n$ that gives minimum value of free energy function, $F(x)$. Moreover, the thermodynamics of protein stability is modeled reasonably well by the Energy Landscape Theory. Stability of medium depends upon the ground and excited states of atoms and nuclear particles which are the simplest as well as basic models while the understanding of complicated system like protein molecules contains far more complex idea and insight into ground and excited states. The ground state of native structure is well illustrated by using energy landscape model where energy represents a function of the topological alignment of atoms. Energy values generated by mountains and ridges are representation of spatial surface with the large number of different co-ordinates.

Figure 2: Uses of Energy Landscape Models.

Protein Folding and Unfolding

Free and attached ribosomes are involved in the production of proteins within the biological system in the form of linear polypeptide chain. The primary or linear structure is converted into stable three dimensional structures with the help of folding process. The whole process of folding is controlled thermodynamically or kinetically by other proteins and/or enzymes, like molecular chaperones. The process of folding represents that the information required for the precise folding is available in the polypeptide chain. On the other hand, the three dimensional structure of proteins are damaged or affected by various reasons including miscoding due to errors in protein synthesis and/or misfolding due to errors in protein folding.

Molecular Chaperones

The chaperones have significant impact on the process of protein folding. As far as their function is concerned, it belongs to the class of proteins that assist in precise folding of proteins. The molecular chaperones are divided into following classes;

Class I chaperones exhibit the affinity to bind with the hydrophobic regions, subsequently preventing aggregation and unfolding polymer is transported to various organelles.

Class II chaperones present inside the organelles assist in misfolding with the help of multiple bonds. It has been observed that when cells are under stress proteins folding become improper.

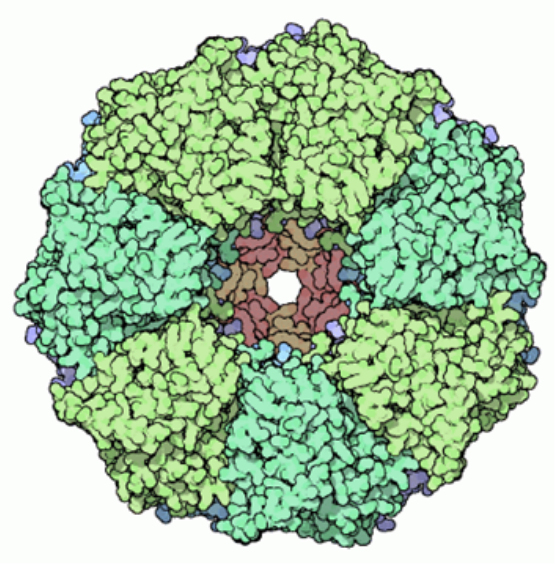


Figure 3: Molecular Chaperone (Top View, Taken From Pdb).

Intramolecular chaperones

Specific sequence of amino acids is essential in the primary structure for proper folding, this sequence is called intra- molecular chaperones and this part is cleaved by cellular proteases and precise folding is accomplished. The stability of a folded protein against denaturation or aggregation is frequently a few times the typical strength of a hydrogen bond in water (About 5~kcal/mol). This stability results from large competing effects that arise from hydration effects and the intra-molecular interactions in the protein. Biological systems also include ions and osmolytes (small organic solutes) in the solvent matrix that can change this delicate balance of interactions for stabilizing a protein. The adaptability in response to various stresses is seen in all living systems. The phenomena underlying such alteration are of fundamental importance in understanding how the solvent controls structure of biomolecules, function, and organization.

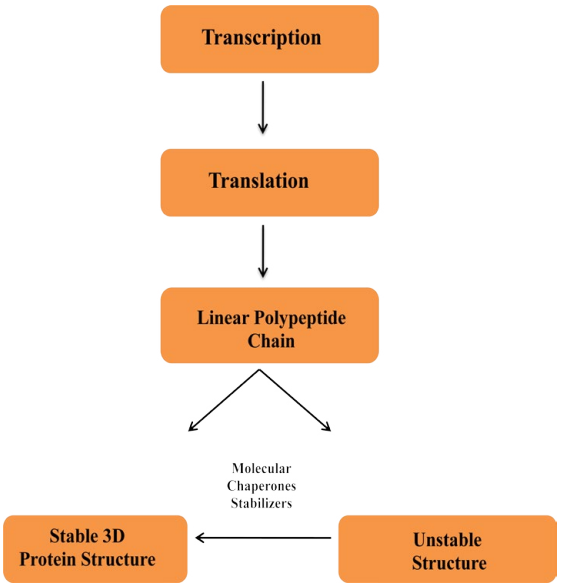


Figure 4: Role of Molecular Chaperones in Protein Stability.

Steric repulsion among the atoms in the covalent bonds exhibit limited elasticity/flexibility in the local region. Unfolded protein has the hydrodynamic characteristics in the presence of denaturants like urea or guanidinium chloride. Experimental investigations support that unfolded proteins are not random coils in a true sense under other physical conditions like pH and temperature extremes in the absence of denaturants. The equilibrium behavior of protein structure is represented by following ways;

N↔U

"N" represents the native (folded) state

"U" represents the unfolded state

The compact intermediate state represents a subset of unfolded state that is altering constantly among the different energetically unfavorable states [2].

Thermodynamics of Unfolding of Azurin

Researchers have investigated the thermodynamics of unfolding of azurin (small blue copper protein) which is responsible for electron

transport in the redox system of certain bacteria. Thermodynamic solidity has been focused in the research of this enzyme. Opposition to the thermal unfolding is remarkable and irreversible unfolding appears at temperatures more than 70°C. This unusual thermal resistance is due to a number of different features like presence of disulphide bridges, hydrophobic effect, intra-molecular hydrogen bonds and stabilization by Cu^{2+} binding. The denaturing and/or unfolding depends upon two effects; a reversible endothermic mechanism which involves the destruction of three-dimensional structure of azurin and an irreversible exothermic process which involves the aggregation of polypeptide chain network [9].

Figure 5: 3d Structure of Azurin. (Rendered by Chimera, ver. 1.8)

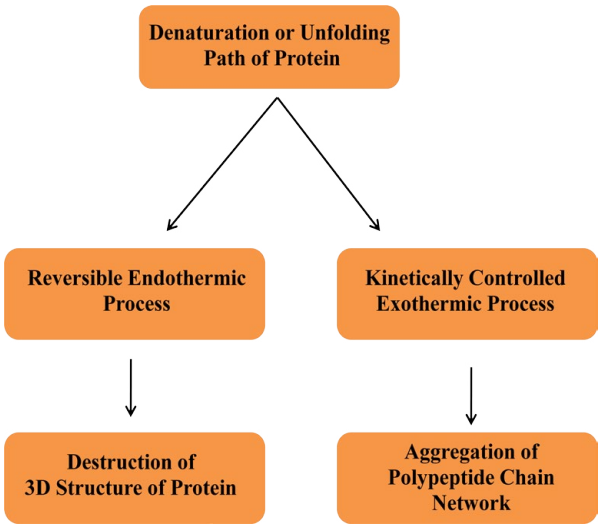


Figure 6: Path Representing The Denaturation Of Proteins.

Pother and Potherse: A Thermodynamic Approach

Researchers have proposed that a protein is not a global thermodynamic system but a complicated one consisting of many subsystems. Scrutinizing the thermodynamic aspects of proteins, two basic concepts of protein dynamics have been introduced by the investigators, there were the pother and the potherse. The pother is considered the basic units of thermal motion of atoms while the potherse, in contrast, is considered as a thermal system comprising many different pothers [10]. Thermodynamic conventional ideas contain particle based concept in the axiomatic theory. On the other hand, the current approach carries pother as the elementary concept of the thermal motion. Different types of thermal motions could be observed in protein including vibrational and rotational in nature due to the side chains of the amino acids [11].

The pother is rotation around the C-N or C-C bond of peptide plain because these are responsible for the overall protein conformational changes. The potherse which contain many subunits/subsystems considered to be thermodynamically stable when the pothers are stable with respect to their thermal motions. As a result, a protein is not a global thermodynamic system but a complex of many subsystems. Moreover, the relationship between pother and potherse could be considered as protein thermodynamic structure. In addition to the 3-D structure, a protein can recognize its thermodynamic structure under different conditions [12,13].

The three-dimensional structure of protein molecules holds the inner thermal motion of amino acids and vice versa. Thermal motions of individual particles influence the overall native form of protein structure. Therefore, the potherse has features of both protein dynamics and stability of protein structure.

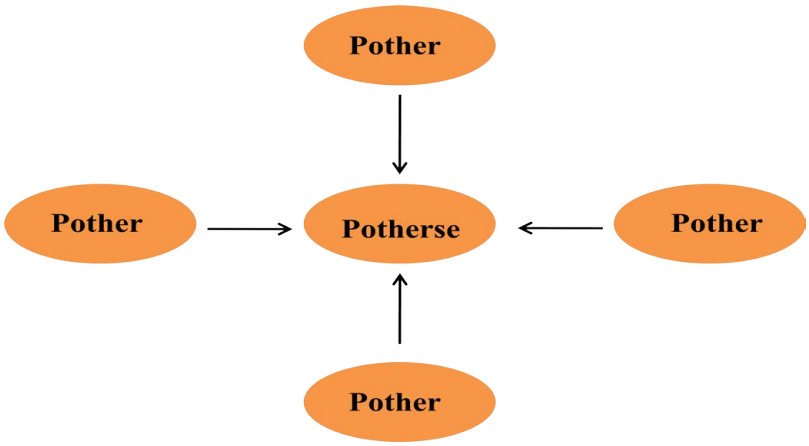


Figure 7: A Potherse is Consists of Many Pothers.

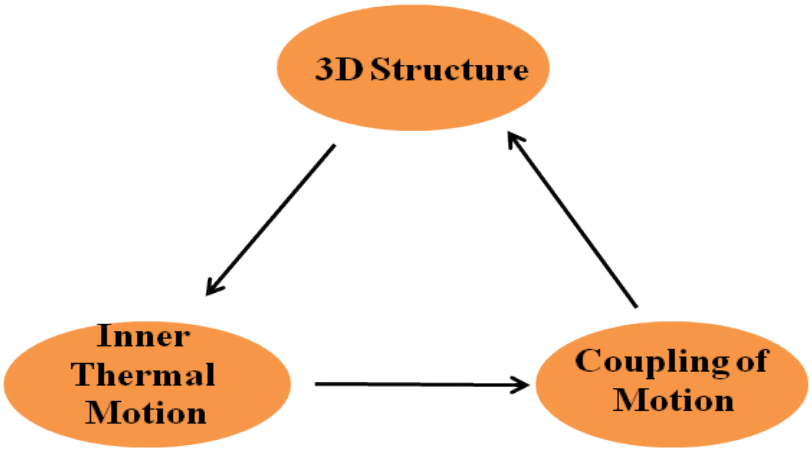


Figure 8: 3-D Formation of Protein within a Potherse.

Cells are the basic units of life and they work as a unit with internal as well as external environment and the molecular mechanisms involved in cell physiology are valuable for understanding in the biological sciences. These tiny structures are incredibly complicated, self-sufficient, microscopic thermodynamic units/systems. Despite their complexity, they are governed by the same principles that apply to all physicochemical systems. Exclusively, they handle energy in the same way that larger or macroscopic thermodynamic systems do. The basic concepts of thermodynamics are essential to an understanding of the molecular mechanisms underlying all cell functions [14]. Thermodynamic principles are essential to an understanding of the complex fluxes of energy and information compulsory to maintain cells alive. These microscopic thermodynamic systems are non-equilibrium systems at the micron scale that are maintained in steady state conditions by very refined and complicated processes. Thermodynamics is a fundamental discipline that enables us to comprehend how energy is handled by living organisms. Numerous of the perceptions and concepts are better understood by considering that the laws of thermodynamics are based on the random, linear or non-linear behavior of large sets of molecules.

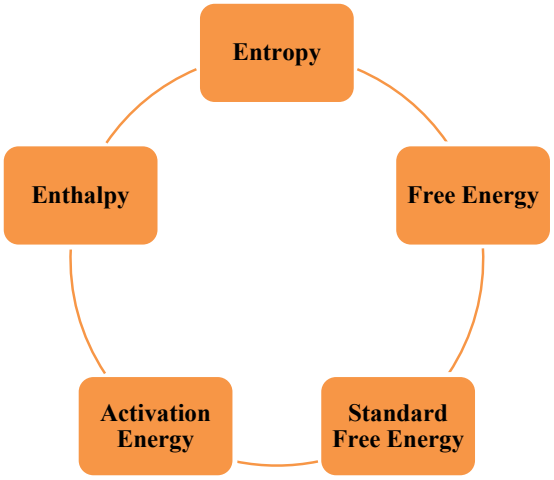


Figure 9: The Basic Concepts of Thermodynamic System.

A qualitative understanding is sufficient for achieving the idea of thermodynamic system. Cells are non-equilibrium systems in which information plays an essential role.

Folding of Proteins and Metalloproteins

Metalloproteins design necessitates an accurately folded protein scaffold containing the suitable number and type of ligands with the accurate geometry to encapsulate and activate the metal for chemical catalysis.

Thermodynamics of Folding of Ribonucleases

Ribonucleases contain three small, extracellular enzymes (Sa, Sa2, and Sa3) produced by different strains of *Streptomyces aureofaciens* with amino acid sequences that are 50% of homology. The unfolding of these enzymes by heat and urea has been studied to determine the conformational stability and its dependence on temperature, pH, NaCl, and the disulfide bond. Urea and guanidinium chloride are responsible for denaturing the protein structure. Both compounds show their interaction in the medium but are weak in their interaction. Molar amount is required for denaturation process. Many ideas have proposed that denaturants show their interaction by direct binding to the protein molecules or by altering the properties of the solvent. On the other hand, disulfide bonds increase the conformational stability of a protein mainly by constraining the unfolded structures or conformations of the protein and by this means decreasing their conformational **Entropy**. Accordingly, the hydrophobic effect and hydrogen bonding make large but equal contributions to the conformational stability of the proteins [15].

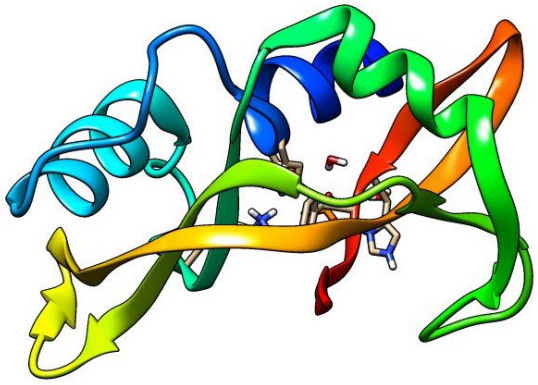


Figure 10: 3D Structure of Ribonuclease A. (Rendered by Chimera, ver. 1.8)

Protein-DNA Interactions and Thermodynamic

In protein-protein interactions, evolutionary conserved amino acids have been studied by many researchers and have been found to be correlated with hotspot residues. A hotspot is a residue whose mutation or change could to be a drop of more than 2kcal/mol in binding free energy [16]. The regions where the hotspot are tightly packed, called Hotspot Regions. Hence, they exhibit rigid packing, they are considered as important role player in stabilization of the structure. Moreover, the contribution of hotspot residues is independent between hotspot regions [17,18]. Various transcription factor proteins recognize and bind with DNA sequences with various affinities [19,20]. The binding ability makes the cells capable of scheming thousands of genes with reasonably few regulatory proteins [21].

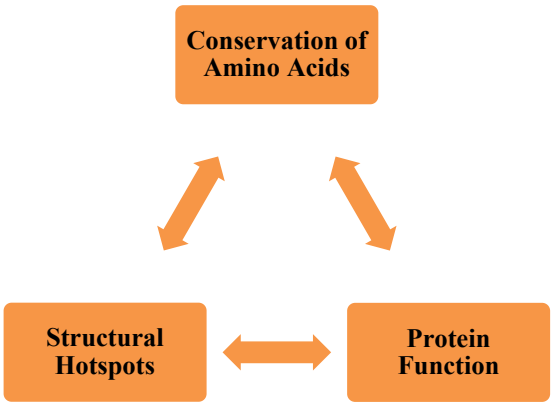


Figure 11: The Relationship between Protein and Other Types of Molecules.

Amino acid residues participate as important components in protein function and are often conserved. Thermodynamic and structural data of protein-DNA interactions is investigated and found a relationship between sequence conservation, free energy and structural sequence. Nearly all stabilizing residues or recognized hotspots are those which occur as clusters of conserved molecules/residues. The compact packing density of the clusters and accessible experimental thermodynamic data of mutations suggest cooperativity between conserved residues in the clusters. Conserved single residue contributes to the strength of protein-DNA complexes to a lesser extent. In protein-protein interactions, conserved residues are highly associated with experimental residue hotspots, giving dominantly and often cooperatively to the stability of protein-protein complexes. On the whole, the conservation models of the stabilizing residues in DNA-binding proteins also emphasize the implication of clustering as compared to single residue conservation.

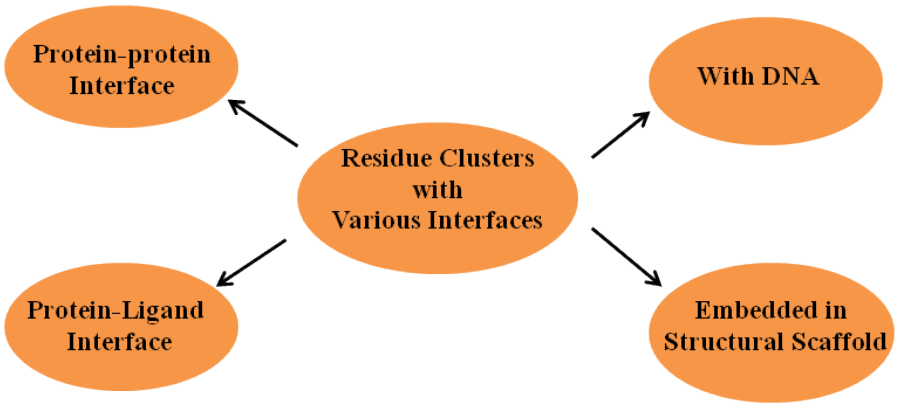


Figure 12: Residue Clusters with Multiple Interfaces.

Binding Sites between Proteins and Other Macromolecules

Interactions between proteins and other biological macromolecules are basic requirement for the variety of mechanisms in a biological system. Studies have revealed that macromolecules contains special site in their structure for the binding with other and/or similar macromolecules. The binding sites for DNA, RNA, obligate protein and non-obligate protein binding and interaction have been studied. Fundamental principles for binding with respect to thermodynamic which govern the interaction of different macromolecules responsible for the chemical reactions at the binding site. Moreover amino acids involved in the binding attribute make a net constructive **enthalpy** and entropic role to the free binding energy [22,23].

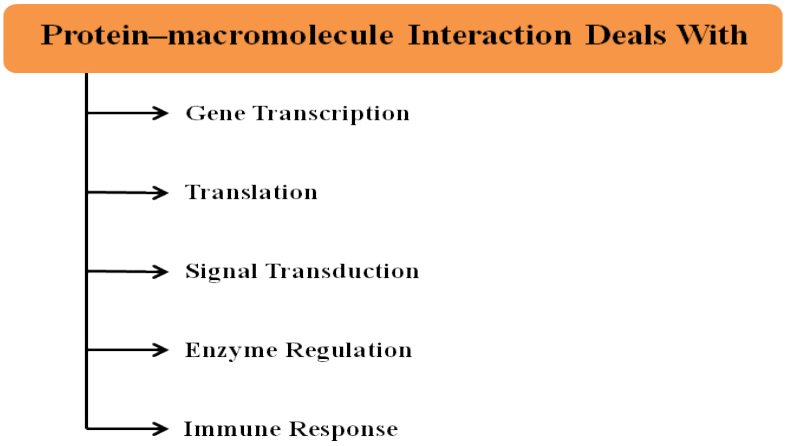


Figure 13: Interactions among protein and other biological macromolecules leads to various cellular mechanisms at molecular level.

Regarding DNA-binding, proteins achieve binding through:

(i) Affinity between the protein and DNA could be observed by the positively charged residues (Arg and Lys) of proteins and negatively charged phosphate backbone of the DNA double helix.

(ii) Specificity could be analyzed by observing side by side van der Waals (VDW) interactions and hydrogen bonding among the positively and negatively charged binding sites [24].

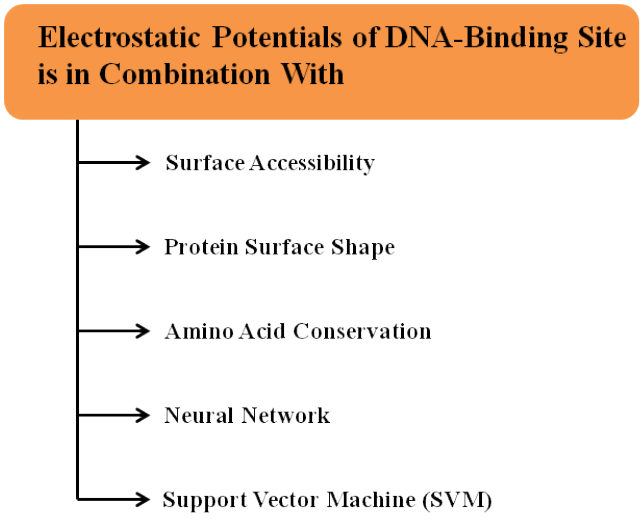


Figure 14: Electrostatic potentials of DNA-binding Site deals with various parameters.

The Hydrophobic effect

The hydrophobic effect is an effective force that acts to minimize the amount of surface area that non-polar molecules expose to their aqueous solvent. The effect originates from the robustly favorable hydrogen-bonding interactions between water molecules that are disrupted by the insertion of a non-polar solute into the solution. The exact dependencies of the free energy of solvation on uncovered surface area, however, depend on the size of the solute. Small solutes can be house without hydrogen-bond breakage through the ordering of the bond network in the surrounding water molecules. On the other hand, for larger solutes the breakage of hydrogen bonds at the surface of the solute is inescapable [25]. Proteins are polymers consisting of chains of amino acids. Each protein folds into an exacting conformation called the native structure, which determines its function, under the physiological condition. The native structure is composed of characteristic local structures called α-helix and β-sheet, hydrogen bonds help for their stabilization. It has been proven experimentally that the native structure is a thermodynamically stable state, called Anfinsen's dogma.

Computational Approach-Computer Simulation of Proteins for Thermodynamics and Structure Prediction

Computer simulations have developed into an increasingly important tool to study proteins. They are used for regularly complement experimental investigations and often are the only tool to explore processes in the cell. Here is a summary of thermodynamic databases and simulation techniques that are used to study the multiple aspects of proteins regarding thermodynamics.

Sr.#	TOOLS	PARTICULARS
1	**ProTherm**	Thermodynamic database for proteins
2	**ProTherm, 2.0**	Thermodynamic Database for Proteins and Mutants
3	**ProTherm, 3.0**	Thermodynamic Database for Proteins and Mutants
8	**Protein Folding Database (PFD)**	A Database for The Investigation of Protein Folding Kinetics and Stability
9	**Protein Folding Database (PFD 2.0)**	An Online Environment for The International Foldeomics Consortium
13	**SRide**	A Server for Identifying Stabilizing Residues In Proteins
16	**FastCo ntact**	A Free Energy Scoring Tool for Protein–Protein Complex Structures
17	**MoViES**	Molecular Vibrations Evaluation Server For Analysis of Fluctuational Dynamics of Proteins and Nucleic Acids

ProTherm

ProTherm has many tools for the users and provide many functions for the determination of various aspects of proteins and their residues and facilitates the users with several options like empirical energy function, distance and contact potential, average assignment analysis, amino acid properties and many other tools to find out the relationship between protein and their stability with respect to the environment in biological system [26].

ProTherm 2.0 is the second version of the database, which not only contains the thermodynamic data of the protein but also the information related to the mutants. Similar its first version, it provides experimental conditions, literature and structural as well as functional information. Protein Mutant Database (PMD) has been developed to facilitate with natural and artificial mutants of proteins.

ProTherm 3.0 contains more features than the previous two versions of the thermodynamic database. As compared to the previous version, it covers more than 10,000 entries for a number of thermodynamic parameters. The protein recognition code from the SWISS-PROT and Protein Data Bank (PDB) has been added into new version and provides cross-links with other databases.

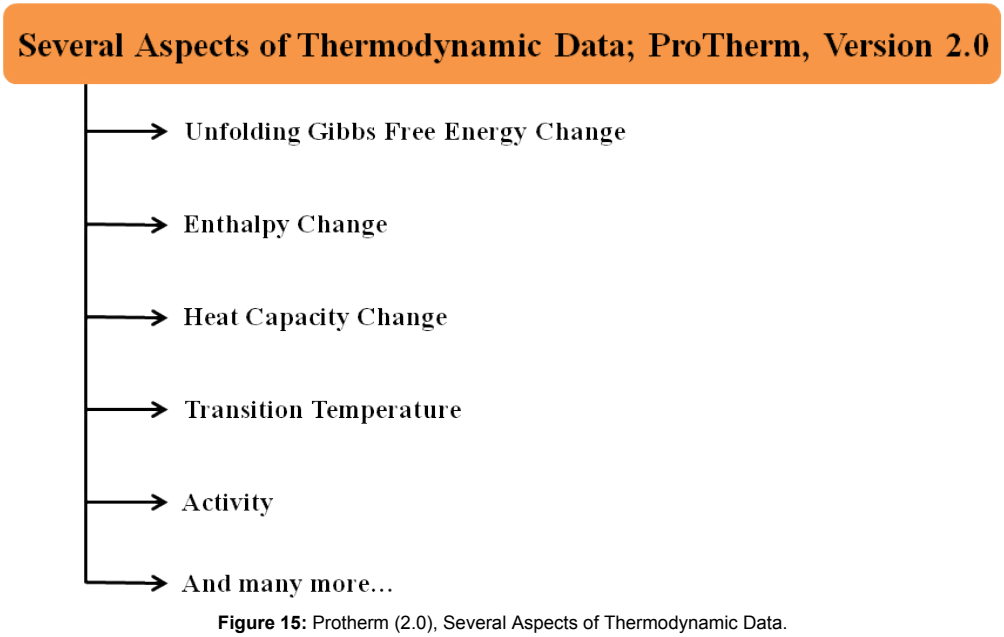


Figure 15: Protherm (2.0), Several Aspects of Thermodynamic Data.

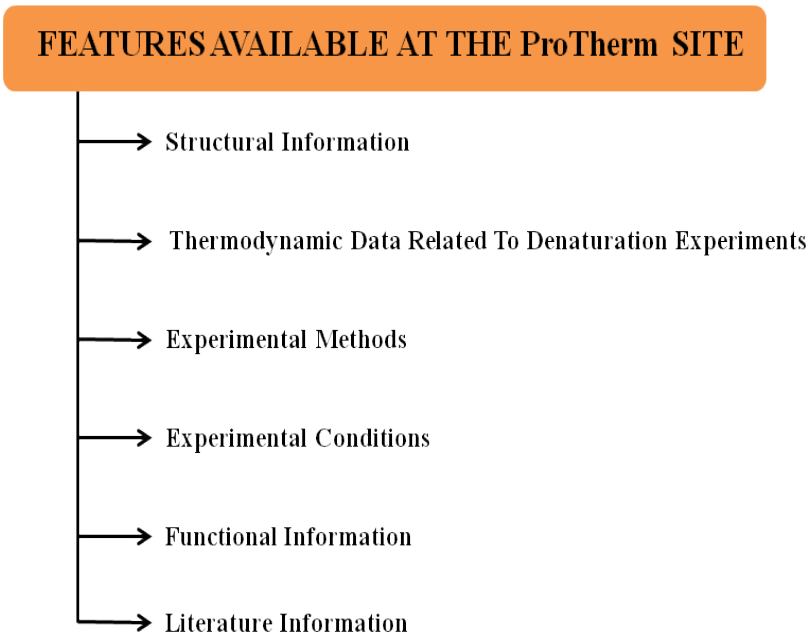


Figure 16: Various Features at the Protherm Site.

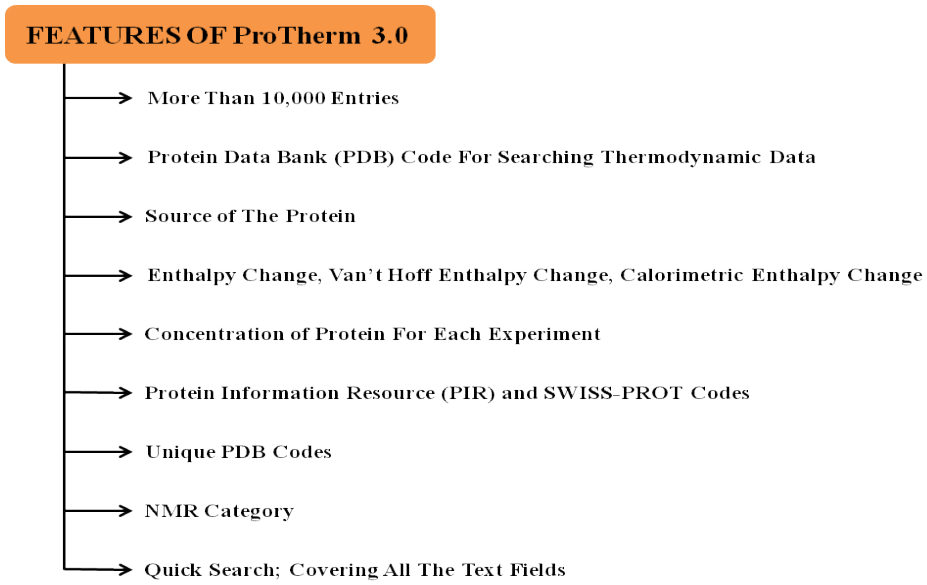


Figure 17: Various features of Protherm Version 3.0. Figure 18: List of Important Information for Protein Folding Experiment at Protein Folding Database (Pfd).

Protein Folding Database (Pfd)

It is really an informative, easily accessible and innovative resource and database that facilitates protein folding data of more than 50 different proteins belongs to various families. Data of structural, kinetic as well as thermodynamic in nature is available on protein folding database that could be accessed through web platform. With the various useful features it links with the other protein databases as well [27].

Protein Folding Database 2.0 provides freely accessible and centralized data repository. It has been built as a comparative repository providing kinetic and thermodynamic features for the folding of protein in a useful and user friendly way [27].

Site Directed Mutator (Sdm)

SDM is a server for predicting effects of mutations on protein stability and malfunction. Site Directed Mutator is a statistical potential energy function that exercises environment-specific amino-acid replacement frequencies within homologous protein families to compute a stability score, which is corresponding to the free energy difference between the wild-type and mutant protein [28].

SRide

SRide server provides data that could be employed for the calculation of van der Waals atomic radius among the amino acids. The residues present in the polypeptide and play a crucial role for the stability of protein molecules are called stabilizing residues (SRs). These residues are selected by using various methods like interaction of a given residue with its neighboring residues and observing the evolutionary conservation in the protein structures [29].

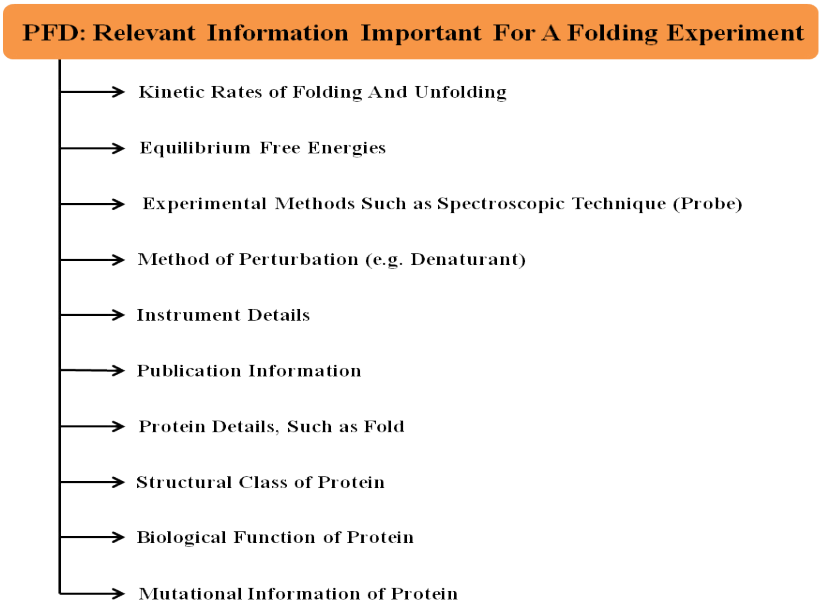


Figure 18: List of Important Information for Protein Folding Experiment at Protein Folding Database (Pfd).

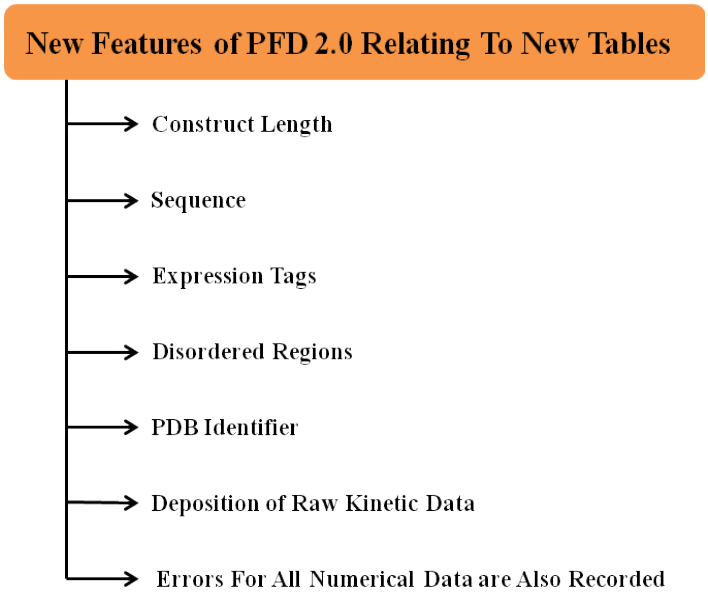


Figure 19: New Features of Pfd, 2.0.

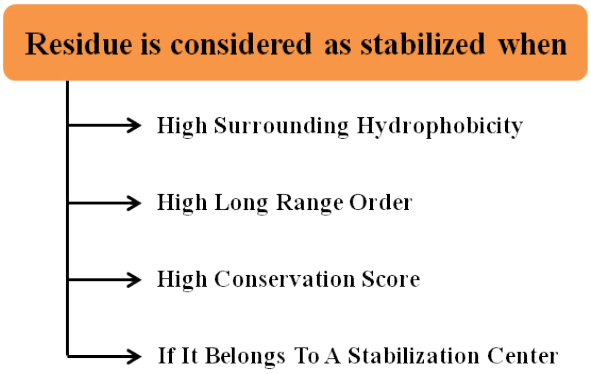


Figure 20: Stability of residue depends upon above mentioned parameters.

As far as interactions are concerned among the neighboring residues, the hydrophobic interaction is the most important one which is assumed to be the primary element for the stability as well as folding of protein molecules. For the stability of the molecule against the unfolding non-covalent and long range interactions are required. It has been investigated that SRs are found in respect to high conservation among the protein molecules.

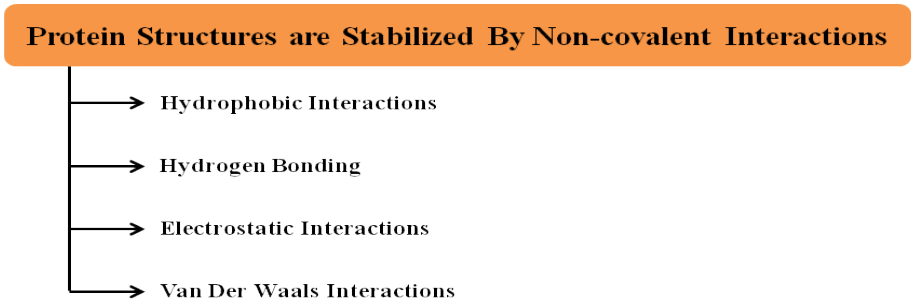


Figure 21: Stability of protein structures and non-covalent interactions.

Fast contact

It is a web server that approximates the straight electrostatic and desolvation interaction free energy between two proteins in units of kcal/mol. The server has been effectively tested and validated, blind sets of docking decoys and scoring refined complex structures, in addition to established helpful predicting protein interactions. FastContact offers exclusive potentials from biophysical insights to scoring and identifying imperative contacts [30].

Brief description of the Algorithm

The code of the server has been built in Fortran 77 while the server itself has been designed in PHP. FastContact performs a speedy computational estimate of the binding free energy between two proteins based on atomic pair wise interactions:

(i) Electrostatic interaction

(ii) Desolvation free energy

(iii) Van-der Waals (VDW) interaction

MoViES

Investigation of vibrational and thermal motions as well as fluctuation at molecular levels are considered to be the valuable tool for studying the structural, dynamic and functional attributes of biological macromolecules including proteins. Freely reachable web program has been developed for the study of vibrational and thermal fluctuations in the biological macromolecules. Vibrational normal mode (VNM) and thermal motion are computed by AMBER molecular mechanics and harmonic approximation method respectively. MoViES (Molecular Vibrations Evaluation Server) have been developed for the study of vibrational dynamics for protein structure and nucleic acids [31].

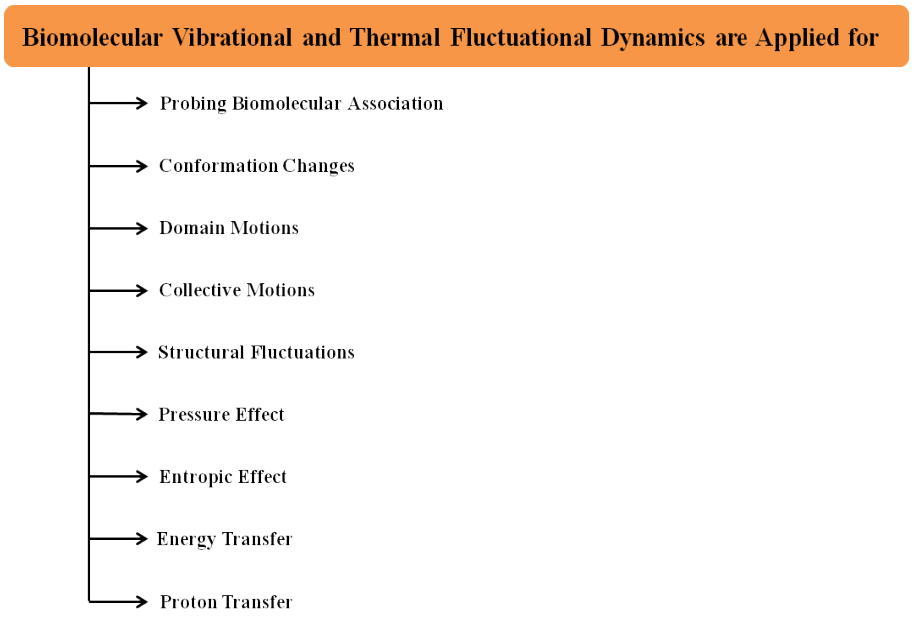


Figure 22: Dynamics related to biomolecular vibrational and thermal fluctuations.

References

1. Richards FM (1991) The Protein Folding Problem. Scientific American, Nature Publishing Group, USA.
2. Creighton TE (1990) Protein folding. Biochem J 270: 1-16.
3. Frauenfelder H, Wolynes PG (1994) Biomolecules: Where the Physics of Complexity and Simplicity meet. Physics Today 47: 58-64.
4. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254: 1598-1603.
5. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21: 167-195.

6. Shea JE, Onuchic JN, Brooks CL 3rd (1999) Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. Proc Natl Acad Sci U S A 96: 12512-12517.

7. Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. Nature 450: 964-972.

8. Dill KA (1999) Polymer principles and protein folding. Protein Sci 8: 1166-1180.

9. La-Rosa C, Milardi D, Grasso D, Guzzi R, Sportelli L (1995) Thermodynamics of the Thermal Unfolding of Azurin. The Journal of Physical Chemistry 99: 14864-14870.

10. Zhao Q (2001) Irreversible thermodynamics theory of protein folding and protein thermodynamics structure. Prog Biochem Biophys 28: 429-435.

11. Benkovic SJ, Hammes-Schiffer S (2003) A perspective on enzyme catalysis. Science 301: 1196-1202.

12. Jiang Y, Lee A, Chen J, Ruta V, Cadene M, et al. (2003) X-ray structure of a voltage-dependent K+ channel. Nature 423: 33-41.

13. Sagermann M, Gay L, Matthews BW (2003) Long-distance conformational changes in a protein engineered by modulated sequence duplication. Proc Natl Acad Sci U S A 100: 9191-9195.

14. Klymkowsky MW (2010) Thinking about the conceptual foundations of the biological sciences. CBE Life Sci Educ 9: 405-407.

15. Pace CN, Hebert EJ, Shaw KL, Schell D, Both V, et al. (1998) Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. J Mol Biol 279: 271-286.

16. George RA, Spriggs RV, Bartlett GJ, Gutteridge A, MacArthur MW, et al. (2005) Effective function annotation through catalytic residue conservation. Proc Natl Acad Sci U S A 102: 12299-12304.

17. Keskin O, Ma B, Nussinov R (2005) Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol 345: 1281-1294.

18. Li X, Keskin O, Ma B, Nussinov R, Liang J (2004) Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. J Mol Biol 344: 781-795.

19. Stormo GD, Fields DS (1998) Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci 23: 109-113.

20. Wang X, Gao H, Shen Y, Weinstock GM, Zhou J, et al. (2008) A high-throughput percentage-of-binding strategy to measure binding energies in DNA-protein interactions: application to genome-scale site discovery. Nucleic Acids Res 36: 4863-4871.

21. Bulyk ML (2003) Computational prediction of transcription-factor binding site locations. Genome Biol 5: 201.

22. Brenner SE, Levitt M (2000) Expectations from structural genomics. Protein Sci 9: 197-200.

23. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, et al. (2002) Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci U S A 99: 11664-11669.

24. Chen YC, Lim C (2008) Common physical basis of macromolecule-binding sites in proteins. Nucleic Acids Res 36: 7078-7087.

25. Chandler D (2005) Interfaces and the driving force of hydrophobic assembly. Nature 437: 640-647.

26. Gromiha MM, Sarai A (2010) Thermodynamic database for proteins: features and applications. Methods Mol Biol 609: 97-112.

27. Fulton KF, Devlin GL, Jodun RA, Silvestri L, Bottomley SP, et al. (2005) PFD: a database for the investigation of protein folding kinetics and stability. Nucleic Acids Res 33: 279-283.

28. Worth CL, Preissner R, Blundell TL (2011) SDM--a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39: 215-222.

29. Magyar C, Gromiha MM, Pujadas G, Tusnády GE, Simon I (2005) SRide: a server for identifying stabilizing residues in proteins. Nucleic Acids Res 33: 303-305.

30. Champ PC, Camacho CJ (2007) FastContact: a free energy scoring tool for protein-protein complex structures. Nucleic Acids Res 35: 556-560.

31. Cao ZW, Xue Y, Han LY, Xie B, Zhou H, et al. (2004) MoViES: molecular vibrations evaluation server for analysis of fluctuational dynamics of proteins and nucleic acids. Nucleic Acids Res 32: 679-685.

Advances in Protein Chemistry

Chapter: Role of Matrix Metalloproteinases in Cancer

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published September, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Role of Matrix Metalloproteinases in Cancer

Mohd. Afaque Ansari[1], Sibhghatulla Shaikh[1], Ghazala Muteeb[2], Syed Mohd. Danish Rizvi[1]*, Shazi Shakil[3], Aftab Alam[3], Rashmi Tripathi[1], Fauzia Ghazal[1], Deboshree Biswas[1], Mohammad Akram[4], Ajijur Rehman[1], Saeedut Zafar Ali[5], Anil Kumar Pandey[3] and Ghulam Md Ashraf[6]

[1]Department of Biosciences, Integral University, Lucknow, India

[2]Freelance researcher, Abu Dhabi, UAE

[3]Department of Bio-engineering, Integral University, Lucknow, India

[4]Department of Biology, Rabigh College of Science and Arts, King Abdulaziz University, Jeddah, Saudi Arabia-21589

[5]Interdisciplinary Biotechnology Unit, Aligarh Muslim University, Aligarh, India

[6]King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia

***Corresponding author:** Dr. Syed Mohd. Danish Rizvi, Assistant Professor, Department of Biosciences, Integral University, Lucknow 226026, India, Tel: +91-8004702899; Fax: +91-522-2890809; E-mail: syedrizvi10@yahoo.com, shazibiotech@gmail.com

Introduction

Matrix metalloproteinases (MMPs) are endopeptidases which depend on zinc for their activity. They belong to the metzincin family of enzymes that codes for a highly conserved zinc binding motif. MMPs have a role in degrading all kinds of extracellular matrix proteins and can also act upon a number of bioactive molecules. These are also called matrixins. MMPs cleave cell surface receptors; they release apoptotic ligands like FAS ligand, and also mediate activation or inactivation of chemokine and cytokine [1]. MMPs effect many behavioral patterns of the cell; like its proliferation, differentiation, migration and apoptosis. In addition, they play a major role in angiogenesis and host defence.

The first MMP was discovered when Jerome Gross and Charles Lapiere (1962) found out enzymatic activity during metamorphosis of the tadpole tail. They observed that the collagen triple helix degraded when the tadpole tail placed in collagen matrix metamorphosed [2]. Their suggestion was that the tadpole tissue produced a protease which was diffusible, but as EDTA could inhibit cleavage, it suggested an involvement of MMP.

Released as an inactive proenzyme, MMPs are activated by factors which are regulated by TIMP (tissue inhibitors of matrix metalloproteinases). The endogenous inhibitor family of TIMP is formed by four enzymes. Brew et al., reported in 2010 that pathological conditions may develop due to imbalance in the levels of MMP and TIMP [3]. Various reports show that an increase in expression of MMPs leads to various inflammatory, malignant and degenerative diseases. Thus there is a possibility that inhibitors of MMPs have therapeutic value [4-6].

Structure

The MMPs have three common domains, namely

• **The pro-peptide**, which is responsible for keeping the enzyme in an inactive form. This domain contains "cysteine switch"- a unique and conserved cysteine residue that interacts with the zinc in the active site. Interaction with the zinc prevents binding and cleavage of the substrate. This domain has to be proteolytically cleaved in order to make the enzyme active. The cleavage is done intracellularly by furin or extracellularly by other MMPs or serine proteinases such as plasmin [7].

• **The catalytic domain**- the structural signature of which is the zinc binding motif. The Zn^{2+} ion, bound by three histidine residues, forms the active site. The active site runs horizontally across the molecule as a shallow groove, and binds the substrate.

• **The hinge region**- A flexible hinge or linker region which connects the catalytic domain to the C-terminal domain. Although this 75 amino acids long region and has no determinable structure, it is very important for the enzyme's stability.

• **The hemopexin-like C-terminal domain**-which is so named due to its sequence similarity to a serum protein, haemopexin. The polypeptide chain of this domain organizes into four β sheets. These β sheets or blades arrange themselves symmetrically around a central channel, resulting into a four-bladed β-propeller structure. The flat surface provided by this structure is believed to be involved in interactions between proteins and is a determinant of substrate specificity, for example TIMP, interacts with this site. However, this domain is not present in plants and nematodes.

ADAM (a disintegrin and metalloproteinase) and ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) are the two families of metzincin proteinases that are closely related to the MMPs. Mostly membrane-anchored and pericellular space functioning ADAMs play roles in fertilization, development, and cancer [8]. ADAMs perform their function in a nonproteolytic manner. The generally secreted and soluble ADAMTS enzymes, function during ECM assembly, ovulation, and cancer, and have a protease, a disintegrin, and one or more thrombospondin domains [9].

MMPs are believed to remodel the ECM as they are capable of degrading ECM molecules. Likewise, MMPs may carry out significant functions during embryonic development as ECM remodelling is considered a critical part of tissue growth and morphogenesis. Additionally, MMPs also influence many cellular functions during development and normal physiology, for example:

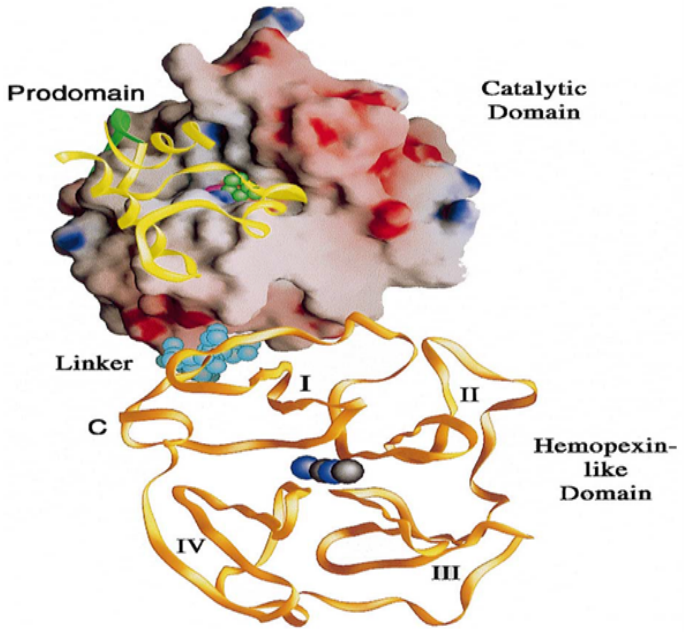


Figure 1: Modelled structure of the full-length pro MMP. The catalytic domain is shown in standard orientation. The pro-domain and the catalytic domain are taken from the experimental proMMP-3 catalytic domain structure, whereas the linker (blue spheres) and the haemopexin-like domain are taken from and attached as seen in the crystal structure of full-length MMP-1. The catalytic domain from the hinge residue, Pro107, onwards is given as a solid surface colored according to the electrostatic surface potential. In the centre of the active-site cleft running from left to right resides the catalytic zinc (pink half sphere, center). The pro-segment is shown as a yellow ribbon with its ordered segments only, that is with the first (only apparently separated) helix, the second helix, the connecting loop, the third helix and the switch loop (running across the catalytic zinc liganding it via the conserved Cys side chain, shown as a green CPK model), and the rocker arm (green color), which swings to the bottom left after activation cleavage. The haemopexin-like domain consisting of the four blades I to IV and viewed at its exit side is shown as a golden ribbon; the single calcium ion (blue sphere, back) located at the entry side (MMP-1) is shown together with the second calcium (central blue sphere) and the two chlorine ions (black spheres) found in the haemopexin-like domains of gelatinase A and MMP-13.

(1) Allowing cell migration through degradation of ECM molecules;
(2) Altering cellular behavior by changing ECM micro-environment;
(3) Modulating the activity of biologically active molecules by direct cleavage, release from bound stores, or the modulating of the activity of their inhibitors.

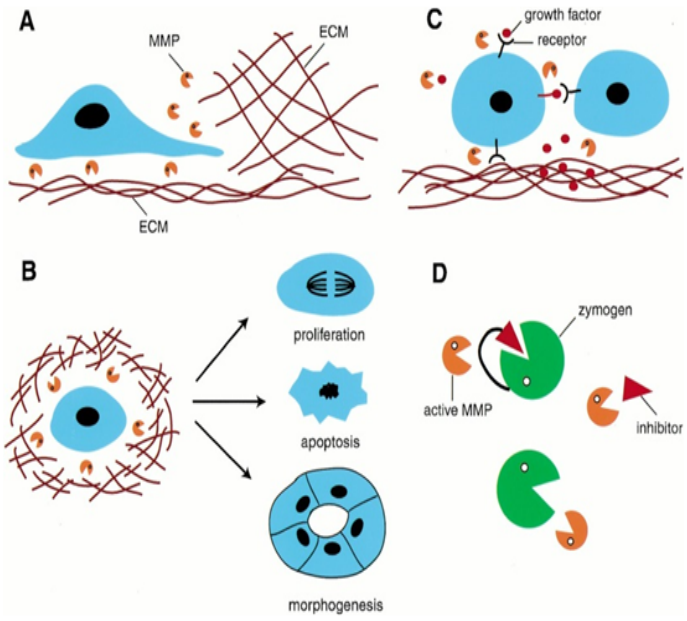


Figure 2: Modes of action of the matrix metalloproteinases. (A) MMPs may affect cell migration by changing the cells from an adhesive to non adhesive phenotype and by degrading the ECM. (B) MMPs may alter ECM microenvironment leading to cell proliferation, apoptosis, or morphogenesis. (C) MMPs may modulate the activity of biologically active molecules such as growth factors or growth factor receptors by cleaving them or releasing them from the ECM. (D) MMPs may alter the balance of protease activity by cleaving the enzymes or their inhibitors.

Classification of MMPs

On the basis of the specificity of MMPs for ECM components, they are divided into collagenases, gelatinases, stromelysins and matrilysins. The common names of the MMPs mirror this classification. Out of eight distinct structural classes of MMPs: five are secreted and three are membrane-type MMPs (MT-MMPs). The MT-MMPs are linked by covalent bonds to the membrane of the cell, the most obvious way of tethering MMP activity to the cell membrane. The other way to localize to the cell surface is by binding to integrins, which is the mode of secreted MMPs or to CD44 or through interactions with cell-surface-associated heparan sulphate proteoglycans, collagen type IV or the extracellular matrix metalloproteinase inducer (EMMPRIN) [7,10-12].

MMP designation	Structural class	Common Name(s)
MMP-1	Simple hemopexin domain	Collagenase-1, interstitial collagenase, fibroblast collagenase, tissue collagenase
MMP-2	Gelatin-binding	Gelatinase A, 72-kDa gelatinase, 72-kDa typeIV collagenase, neutrophil gelatinase
MMP-3	Simple hemopexin domain	Stromelysin-1, transin-1, proteoglycanase, procollagenase activating protein
MMP-7	Minimal domain	Matrilysin, matrin, PUMP1, small uterine metalloproteinase
MMP-8	Simple hemopexin domain	Collagenase-2, neutrophil collagenase, PMN collagenase, granulocyte collagenase
MMP-9	Gelatin-binding	Gelatinase B, 92-kDa gelatinase, 92-kDa type IV collagenase
MMP-10	Simple hemopexin domain	Stromelysin-2, transin-2 MMP-11 Furin-activated and secreted Stromelysin-3
MMP-12	Simple hemopexin domain	Metalloelastase, macrophage elastase, macrophage metalloelastase
MMP-13	Simple hemopexin domain	Collagenase-3
MMP-14	Transmembrane	MT1-MMP, MT-MMP1
MMP-15	Transmembrane	MT2-MMP, MT-MMP2
MMP-16	Transmembrane	MT3-MMP, MT-MMP3
MMP-17	GPI-linked	MT4-MMP, MT-MMP4
MMP-18	Simple hemopexin domain	Collagenase-4 (*Xenopus*; no human homologue known)
MMP-19	Simple hemopexin domain	RASI-1, MMP-18‡
MMP-20	Simple hemopexin domain	Enamelysin
MMP-21§	Vitronectin-like insert	Homologue of *Xenopus* XMMP
MMP-22	Simple hemopexin domain	CMMP (chicken; no human homologue known)
MMP-23 ‖	Type II transmembrane	Cysteine array MMP (CA-MMP), femalysin, MIFR,MMP-21/MMP-22¶
MMP-24	Transmembrane	MT5-MMP, MT-MMP5
MMP-25	GPI-linked	MT6-MMP, MT-MMP6, leukolysin
MMP-26	Minimal domain	Endometase, matrilysin-2
MMP-27#	Simple hemopexin domain	
MMP-28	Furin-activated and secreted	Epilysin
No designation	Simple hemopexin domain	Mcol-A (Mouse)
No designation	Simple hemopexin domain	Mcol-B (Mouse)
No designation	Gelatin-binding	75-kDa gelatinase (chicken)

Table 1: The Matrix Metalloproteinases Family.

*MMP-4, -5 and -6 have been abandoned. ‡When MMP-19 was cloned it was initially called MMP-18. However, an MMP from *Xenopus* had already received that designation, and therefore this MMP is now known as MMP-19. §The cloning of a partial fragment of human MMP-21 has been described, but the sequence has not been submitted to GenBank and the human enzyme has not been characterized. ‖By similarity with mouse and rat MMP-23. ¶Gururajan and colleagues identified two new MMP genes, which they called MMP21 and MMP22. The nucleotide sequences of the two genes are almost identical, so they are now designated MMP23A and MMP23B. #Sequence submitted to GenBank (access no. AF195192). GPI, glycosylphosphatidylinositol; MMP, matrix metalloproteinase; MT-MMP, membrane type MMP; PMN, polymorphonuclear neutrophil; PUMP, putative metalloproteinase

Role of Matrix Metalloproteinase

MMPs1 (Interstitial collagenase)

Collagen type I, a major component of bone ECM is degraded by Matrix Metalloproteinase 1 (MMP-1). MMP-1 was significantly down-regulated, while TIMP-1 levels were increased, in a time- and pressure-dependent manner in a smooth muscle cell (SMC) mechanical strain model. Fibroblasts, keratinocytes, endothelial cells, monocytes and macrophages express MMP-1. Additionally, a misexpression screen set up to identify molecules required for motoneuron development also resulted in isolation of *Mmp1*. MMP-1 encoding mRNA was expressed at considerably higher levels in Human OS cells in primary culture than normal human bone cells.

MMPs2 (Gelatinase-A, 72 kDa gelatinase)

Whole-mount RNA in *in situ* hybridization characterized the expression pattern of *Mmp2*. Expression of *Mmp2* widely takes place in the embryonic CNS, which is in contrast to *Mmp1*. Expression of MMP-2 and beta-catenin loss have a role in the pathogenesis and progression of ESC. Recently, it has been shown that DNAzyme generated against MMP-2 mRNA reduced the expression of the enzymes *in vitro*, and the size of the C6-glioma *in vivo*, in the animal model. An important role is played by decreased E-cadherin in the development of both ESC and EEC.

MMPs3 (Stromelysin 1)

The pattern of IGFBP-3 degradation products produced by MMP-3 is identical in size to that produced by pregnancy serum. The stromelysin subgroup contains stromelysin-1 (MMP-3). A series of apoE/MMP double knockout mice were used in recent studies on atherosclerotic plaque stability to indicate that MMP-3 limits plaque growth and enhances plaque stability, and thus plays a protective role.

MMPs7 (Matrilysin, PUMP 1)

A big role in the invasion and metastasis of cancer is played by Matrix metalloproteinase-7 (MMP-7), the matrix-degrading enzyme. Studies have shown that oligonucleotides antisense to MMP-7 inhibit the higher rate of spreading of gastric gland cells infected with *H. pylori* cultures. MMP-7 mRNA was expressed in 53% of primary gastric cancers, but not in normal gastric mucosa, fibroblasts, or mesothelial cells. Induction of MMP-7 takes place during the response of epithelial cell to bacterial infection.

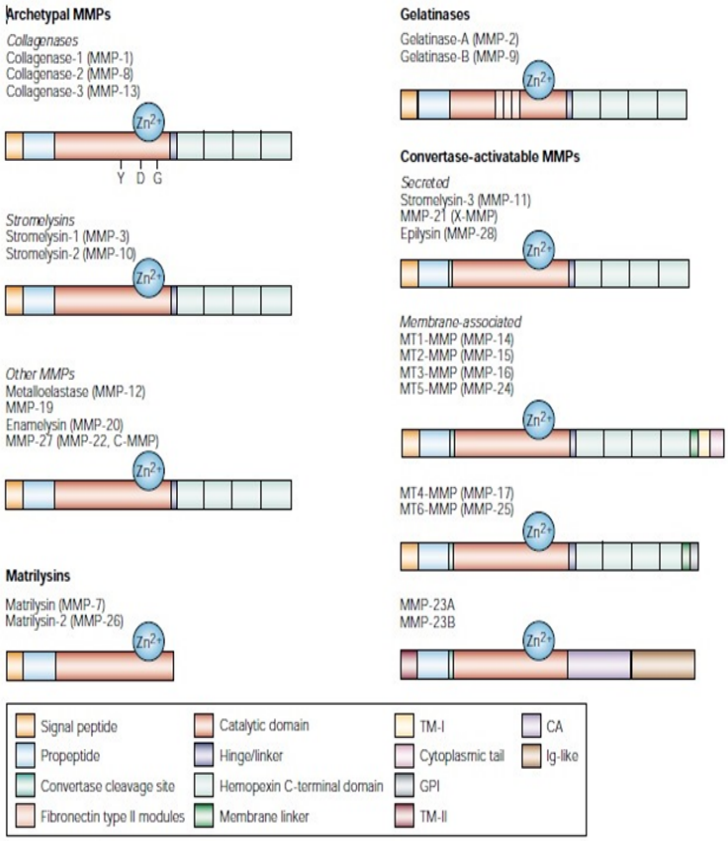


Figure 3: Human MMPs. Schematic representation of the structure of the 24 human matrix metalloproteinases (MMPs), which are classified into four different groups on the basis of domain organization. Archetypal MMPs contain a signal peptide (necessary for secretion), propeptide, a catalytic domain that binds zinc (Zn^{2+}) and a hemopexin carboxy (C)-terminal domain. Y, D, and G represent tyrosine, aspartic acid and glycine amino acids that are present in the catalytic domain of all collagenases. Matrilysins contain the minimal domain organization that is required for secretion, latency and catalytic activity. Gelatinases contain fibronectin type II modules that improve collagen and gelatin degradation efficiency. Convertase activatable MMPs contain a basic insert in the propeptide that is targeted by furin-like proteases (convertase cleavage site). MMPs that belong to this group can be secreted enzymes, or membrane-anchored via GPI (glycosylphosphatidylinositol), type I or type II transmembrane (TM) segments. MMP-23A and MMP-23B contain unique cysteine array (CA) and immunoglobulin (Ig)-like domains in their C-terminal region.

MMPs8 (Neutrophil collagenase)

Neutrophil collagenase, a collagen cleaving enzyme, is present in the connective tissue of most mammals. It is also known as (MMP-8) or PMNL collagenase (MNL-CL). It has an exclusive pattern of expression in inflammatory conditions, and is therefore unique among the family of matrix metalloproteinases (MMPs). MMP-8 mRNA and protein were expressed in all the 3 cell types of human atheroma *in situ*.

MMPs9 (Gelatinase -B, 92 kDa gelatinase)

It is also known as 92 kDa type IV collagenase, 92 kDa gelatinase or gelatinase B (GELB) [13]. MMP-9 releases skit and this enables translocation of BM repopulating cells to a permissive vascular niche, which favours differentiation and reconstitution of the stem/progenitor cell pool.

MMPs10 (Stromelysin 2)

MMP10 gene in humans encodes Stromelysin-2 enzyme, which is also known as matrix metalloproteinase-10 (MMP-10) or transin-2 [14,15].

MMPs11 (Stromelysin-3)

In humans, *MMP11* gene encodes Stromelysin-3 (SL-3) or (MMP-11) [16-19]. Role of Matrix metalloproteinase-11 in neointima formation was tested with the use of a vascular injury model in wild-type (MMP-11+/+) and MMP-11–deficient (MMP-11–/–) mice. Probably, MMP-11 overexpression was associated with the aggressiveness of ovarian carcinoma.

MMPs12 (Macrophage metalloelastase)

Investigation into the role of MMP-12 in the development of COPD in human smokers was undertaken in animal models, and it suggested a predominant role for MMP-9 and MMP-12 in the pathogenesis of pulmonary inflammation.

MMP13 (Collagenase 3)

MMP13 gene in humans encodes Collagenase 3 enzyme [20,21]. Expressed by chondrocytes and synovial cells in human OA and RA, MMP-13 is thought to play a critical role in cartilage destruction. It has been reported that in MMP-13 KO mice, degradation of connective tissue growth factor in wound tissue was transiently prevented. MMP-13 remains the major MMP expressed by chondrocytes to degrade their matrix, when they are stimulated with retinoic acid.

MMP14 (MT1-MMP)

Membrane type 1-matrix metalloprotease (MT1-MMP or MMP-14) is a major activator of pro-MMP-2 and is essential for skeletal development. It is generated *in vitro* by cleavage of membrane-bound native MT1-MMP with several recombinant MMPs, including both active MT1-MMP and MMP-2.

MMP15 (MT2-MMP)

Ueno et al. found a correlation between positive nodal status and the expression of 15 mRNA. Hypocellular ECs at E10. 5 were displayed by mice with targeted snai1 knockdown. This was associated with decreased expression of mesenchyme cell markers and down regulation of the matrix metalloproteinase (mmp) family member, mmp15.

MMP16 (MT3-MMP)

According to a numerical nomenclature for matrix metalloproteinases, this is the new name for MT3-MMP [Membrane-type matrix metalloproteinase-3]. In end-stage osteoarthritis, MT3-MMP expression is elevated in human cartilage. PDGF and fibronectin can upregulate MMP-16 expression by cultured vascular smooth muscle cells under pathologic conditions.

MMPs17 (MT4-MMP)

MMPs-17 (MT4-MMP) is a member of the MT-MMP subfamily. They are anchored to the plasma membrane via a glycosyl-phosphatidyl inositol (GPI) anchor, which confers these enzymes a unique set of regulatory and functional mechanisms that separates them from the rest of the MMP family.

MMP18 (Collagenase 4, xcol4, Xenopus collagenase)

MMP-18 is expressed in the migrating macrophages, and bands corresponding to mRNA for MMP-18 are present in both HTM and CB tissue.

MMP19 (RASI-1, occasionally referred to as stromelysin-4)

MMP-19 was revealed as a novel mediator in laser capture microscope followed by microarray analysis in hyperplastic epithelial cells adjacent to fibrotic regions. Expressed in human epidermis and endothelial cells, it has roles in cellular proliferation, migration, angiogenesis and adhesion. Yu et al., 2012, identified multiple transcript variants encoding distinct isoforms for this gene [22].

MMP20 (Enamelysin)

An MMP-20 mutation which alters the normal splice pattern and results in premature termination of the encoded protein has been associated with amelogenesis imperfecta.

MMP21 (X-MMP)

MMP-21 enhances tumor invasion and metastasis ability in some solid tumors. MMP-21 expression has been investigated in 296 cases of gastric cancer by immunohistochemistry assay.

MMP23A (CA-MMP)

Unlike other MMPs, MMP23a does not possess the signal sequence. This suggests that it may act intracellularly. MMP-23 has a short prodomain and contains a single cysteine residue that can be part of the cysteine-switch mechanism operating for maintaining enzyme latency.

MMP23B

MMP23B degrades various components of the extracellular matrix. *Mmp23b* was identified as a gene linked to the genomic locus of an enhancer trap transgenic zebrafish line in which GFP expression was restricted to the developing liver.

MMP24 (MT5-MMP)

TIMPs inhibit all MMPs, except MMP -24.

MMP25 (MT6-MMP)

Membrane-type MMPs (MMP -25, also called MT1-, MT2-, MT3-, MT4-, MT5-, and MT6-MMP, respectively) are structurally similar to the other classes of MMPs but are anchored to the exterior of the cell membrane. It is highly expressed in leukocytes and in some cancer tissues.

MMP26 (Matrilysin-2, endometase)

MMP-26 has 998 mRNA nucleotides and no transcript variant. RT-PCR, immunofluorescence analysis and flow cytometry determined the mRNA and protein expression of MMP-26 by. It is the smallest member of the matrix metalloproteinase. . The encoded protein degrades type IV collagen, fibronectin, fibrinogen, casein, vitronectin, alpha 1-antitrypsin, alpha 2-macroglobulin, and insulin-like growth factor-binding protein 1, and activates *MMP9* by cleavage.

MMP27 (MMP-22, C-MMP)

mRNAs for MMP-27 are generally expressed at a lower level.

MMP28 (Epilysin)

Matrix metalloproteinase-28 (MMP-28, epilysin) is highly expressed in the skin by keratinocytes, the developing and regenerating nervous system and a number of other normal human tissues. MMP-28 expression is associated with cell proliferation during epithelial repair and is tightly regulated spatially and temporally during wound repair. In primary keratinocytes, expression of MMP-28 is upregulated by treatment with TNF-alpha [23].

Regulation of MMPs

MMPs are synthesized as inactive zymogens. They are kept inactive by an interaction between a cysteine-sulphydryl group in the propeptide domain and the zinc ion bound to the catalytic domain. Their activation requires proteolytic removal of the propeptide prodomain. Activation of majority of MMPs occurs outside the cell by other activated MMPs or serine proteinases. MMP-11, MMP-28 and MT-MMPs can however also be activated by furin-like serine proteinases present intracellularly before they reach the cell surface [7]. Activation of MMP-2 at the cell surface occurs by a unique multistep pathway involving MMP-14 (MT1- MMP) and the tissue inhibitor of metalloproteinases 2 (TIMP-2) [24]. During this process, TIMP-2 binds MMP-14 at its amino terminus and pro-MMP-2 at its carboxyl terminus, allowing an adjacent, non-inhibited MMP-14 to cleave the bound pro-MMP-2. MMP-14 does not fully activate MMP-2, as activatedMMP-2 s necessary to remove a residual portion of the MMP-2 propeptide [25]. Alternatively, activation of Pro-MMP-2 might also occur by MMP-15 through a mechanism not requiring TIMP-2.

An abundant plasmaprotein in tissue fluids, α2-macroglobulin acts as the main inhibitor of MMPs by forming a α2-macroglobulin–MMP complex which binds to a 'scavenger receptor' and gets irreversibly cleared by endocytosis [26]. The debris of the cell is scavenged by scavenger receptors, which forms a broad class of receptors. They also have other activities, such as adhesion. Similarly, thrombospondin-2 forms a complex with MMP-2, facilitating scavenger-receptor-mediated endocytosis and clearance [27]. Thrombospondin-1 binds to pro-MMP-2 and -9, directly inhibiting their activation [28,29]. TIMPs -1, -2, -3 and -4 remain the best-studied endogenous MMP inhibitors. All of them reversibly inhibit MMPs in a 1:1 stoichiometric fashion [8]. Wang et al., 2000, studied Timp-2-deficient mice and showed that the dominant physiological function of TIMP-2 is activation of MMP-2 [30].

MMP Substrate

MMPs make cell migration easier by degrading the structural components of ECM. Cells, on the other hand, have receptors for structural ECM components for example, integrins, therefore cleavage of ECM proteins by MMPs affects cellular signalling and functions [31]. Cleavage of ECM components by MMPs generates fragments with new functions: cleavage of laminin-5 and collagen type IV results in exposure of 'cryptic sites' that promote migration [32,33]. A part of protein which is normally hidden within its three dimensional structure constitutes the 'Cryptic site'. It might get exposed due to conformational changes in the protein. Additionally, cleavage of insulin-like growth-factor-binding protein (IGF-BP) and perlecan releases IGFs and fibroblast growth factors (FGFs), respectively [34-36]. Moreover, MMPs along with the ADAMS participate in the release of cell-membrane-bound precursor forms of many growth factors, including transforming growth factor-α (TGF-α) [37]. However, TGF-β is released by MMP-2 and MMP-9 from an inactive extracellular complex and made available biologically [38].

MMP-2 cleaves the FGF receptor 1, a growth factor receptor [39]. Two members of the epidermal-growth-factor receptor (EGFR) family — HER2/neu (also known as ERBB2) and HER4 (also known as ERBB4) — and the hepatocyte-growth-factor receptor c-MET are substrates for unidentified MMPs or ADAMs [40-42].

MMPs also act on cell-adhesion molecules. For instance, cleavage of E-cadherin and CD44 results in the release of fragments of the extracellular domains and an increased invasive behaviour [43,44]. Cleavage of the α-v integrin subunit precursor by MMP-14 enhances cancer-cell migration. In addition, MMPs cleave other MMPs and proteinase inhibitors such as serpins [7].

Inappropriate expression of MMP activity results in the pathogenic mechanism associated with a wide range of diseases which include;

✓ Tissue breakdown and remodelling during invasive tumor growth and tumor angiogenesis [45]

✓ Degradation of myelin-basic protein in neuro-inflammatory diseases [46]

✓ Increased matrix turnover in restenotic lesions [47]

✓ Liver fibrosis [48]

✓ Loss of aortic wall strength in aneurysms [49]

✓ Tissue degradation in gastric ulceration [50]

✓ Opening of the blood-brain barrier following brain injury [51]

✓ Breakdown of connective tissue in periodontal disease [52]

✓ Acute lung injury and acute respiratory distress syndrome [53] and,

✓ The destruction of cartilage and bone in rheumatoid and osteoarthritis [54]

However, it is still a subject of interest that which MMPs are involved in which diseases.

MMPs Inhibition and Anticancer Therapy

Synthesis of MMPs is blocked by several agents which prevent them from interacting with the molecules that direct their activities to the cell surface or inhibit their enzymatic activity.

Inhibition of MMP synthesis

MMP synthesis is inhibited directly by transfecting cells with antisense mRNA or oligonucleotides, or by targeting mRNA with RIBOZYMES. In mouse models, this means was used to downregulate MMP7 or 9 to reduced tumour burden or metastasis [55-57]. Indirect methods to reduce MMP expression are inhibition of the signal-transduction pathways that induce MMP transcription. Several drugs in clinical trials inhibit tyrosine kinase receptor signalling and affect MMP expression levels [58]. Halofuginone, a COCCIDIOSTAT that is used in chickens, is a drug regulating MMP gene expression and experimental cancer-cell metastasis [59]. Coccidiostat is a drug used to treat coccidiosis, an intestinal disease that is caused by a protozoan.

Inhibiting interactions between MMP and other proteins

In order to inhibit interaction of MMP with other proteins, MMP-2 is inhibited from binding to αvβ3 integrin. In clinical practice, this type of strategy could be tested by means of specifically targeting cancer-promoting function, and the compound shows promising results in animal experiments [60].

Exploiting MMP activity

Several cytotoxic agents have been developed that are activated by MMPs. This is useful in treatment of tumours. Cytotoxic agents like recombinant proteins containing ANTHRAX TOXIN fused to an MMP cleavage site, are activated by MMP cleavage at the cell surface and are internalized by the cell, followed by cell death [61].

Blocking of MMP

The MMPs are inhibited by specific endogenous (TIMPs), which comprise a family of four protease inhibitors:

- TIMP-1
- TIMP-2
- TIMP-3 and
- TIMP-4

TIMPs might have MMP-independent cancer-promoting activities [62].

Three categories of synthetic MMP inhibitors are:

- The collagen peptidomimetics, which mimic the cleavage sites of MMP substrates. Examples are Batimastat and Marimastat. While Batimastat is no longer tested for the treatment of human cancer, Marimastat has undergone several Phase III clinical trials.

- The collagen non-peptidomimetics, which are synthesized on the basis of the conformation of the MMP active site. Examples include BAY 12-9566, Prinomastat/AG3340, BMS 275291 and CGS 27023A/MMI270. While treatment with BAY 12-9566 in Phase III trials showed poorer survival than for placebo-treated groups, Prinomastat combined with standard chemotherapy did not show beneficial effects compared with chemotherapy alone. Phase II/III clinical trials with BMS 275291 are being recruited [63].

- The tetracycline derivates, which inhibit both the activity and synthesis of MMPs. An example is Col-3 (Metastat) that has entered Phase II trials for Kaposi's sarcoma and advanced brain tumours [64]. A new class of MMP inhibitors are small peptides which can be selected for high specificity for individual MMPs, and one such peptide that inhibits MMP-2 and -9 enzymatic activity shows promising effects in animal experiments [65].

Bisphosphonates, originally developed for the treatment of disturbances in calcium homeostasis and for the prevention of bone metastasis, also inhibit the enzymatic activity of MMPs [66]. Some unconventional MMP inhibitors like AE-941 (Neovastat), an extract from shark cartilage, inhibits MMPs and is now in Phase III clinical trials for the treatment of metastatic non-small-cell lung cancer [67]. A component in green tea being tested in Phase III trial, acts as an MMP-2 and -9 inhibitor *in vitro* [68]. Acetylsalicylic acid reduces the risk of colon cancer, by directly inhibits MMP-2 activity [69].

Therapeutic Inhibition of MMPs

Strategies for blocking MMP gene transcription

General approaches to inhibit MMP gene transcription target extracellular factors signal transduction pathways or nuclear factors that activate expression of these genes. Targeting MMP transcripts using ribozymes or antisense constructs downregulate MMP production by cancer cells [55,57,70].

Extracellular factors

IFN-γ inhibits transcription of several MMPs via the transcription factor STAT1 in diverse human cancer cells [71]. Similarly, IFN-β and IFN-α can also be used for this purpose [72-73]. Blocking of signalling by cytokines or growth factors that upregulate MMPs serves as an alternative approach.

Monoclonal antibodies therapeutically reduce TNF-α-induced MMP production in arthritis, and therefore have a potential in cancer too. In a similar strategy for abolishing MMP production in cancer, blocking of IL-1 or epithelial growth factor (EGF) receptors might be useful [74,75]. Retinoids and TGF-β have been reported to downregulate the expression of MMPs and increase TIMP expression, but other studies have reported the opposite [76-79]. Interestingly, blockade of TGF-β with a soluble TGF-β receptor antagonist inhibits tumour metastasis and the production of active MMP-2 and MMP-9 in mouse models of breast carcinomas [80]. Therefore, proposals to use retinoids or TGF-β as therapeutics to block MMP production in cancer must be reconsidered in light of stimulatory effects of these agents on the production of diverse tumour-associated MMPs. As selecting MMP targets for cancer therapy becomes more complex, the need is to define specific proteases that are involved in a particular tumour at each stage.

Signal transduction

MMP production can also be blocked by targeting the signal-transduction pathways that mediate their induction (Figure 4). Halofuginone (an alkaloid from the medicinal plant *Dichroa febrifuga*) interferes with the TGF-β signalling pathway and inhibits bladder carcinoma metastasis by blocking MMP2 expression [81]. In addition, selective inhibition of p38 MAPK activity with SB203580 abolishes the expression of MMP1, 9 and 13 in transformed keratinocytes and squamous-cell carcinoma cells [82,83]. Inhibitors of other MAPK pathways, including ERK and JNK, also block the production of some MMPs by tumour cells [84]. Malolactomycin D — a potent inhibitor of transcription that is controlled by the RAS responsive element suppresses the expression of several MMPs and the RAS-induced transformed phenotype in NIH3T3 cells, at least in part, by blocking the activation of p38 MAPK [85]. Also, manumycin A — an inhibitor of RAS farnesyltransferase — blocks hyaluronan-mediated MMP-2 secretion in lung carcinoma cells, indicating that RAS signalling is required in this process [86]. Since blocking of general oncogenic signalling pathways-RAS–MAPK, can hamper cancer progression by means of many different mechanisms, therefore it will be crucial to show that therapeutic MMP inhibitors have no deleterious effects on signal transduction.

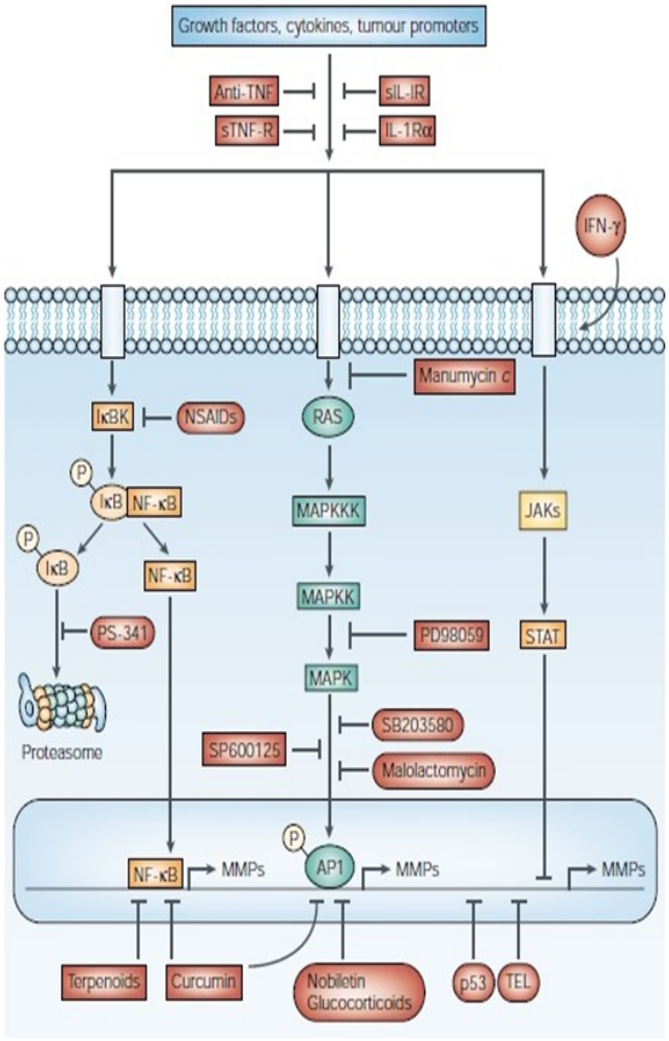


Figure 4: Signalling pathways involved in MMP gene transcription, and potential strategies for therapeutic intervention. Compounds that are able to block the transcription of matrix metalloproteinase (MMP) genes at different levels are shown in red boxes. Extracellular factors, such as interferon-γ (IFN-γ) inhibit MMP transcription via the JAK–STAT signalling pathway. Monoclonal antibodies against tumour necrosis factor-α (anti-TNF), soluble forms of the TNF receptor (sTNF-R), natural antagonists of the interleukin (IL)-1 receptor-α (IL-1Rα) or soluble forms of this receptor (sIL-R) can block signalling pathways initiated by extracellular factors such as TNF-α and IL-1, which induce MMPs in cancer cells. Compounds such as manumycin A, SB203580, malolactomycin, SP600125 or PD98059 act at different levels to block the signal transduction pathways that are associated with MMP transcriptional induction in human tumours. Finally, there are several possibilities to target the nuclear factors that are responsible for MMP transcriptional upregulation. Glucocorticoids, terpenoids, curcumin, nobiletin or NSAIDs (nonsteroidal anti-inflammatory drugs) block the activity of transcription factors such as AP1 and NF-κB, which regulate the transcription of several MMP genes. Similarly, restoring the activity of transcription factors such as p53 and TEL, which negatively regulate MMP expression and the activity of which is lost in human tumours, could downregulate these genes. IFN-γ, interferon-γ; IκB, inhibitor of κB ; IκBK; inhibitor of κB kinase; JAK, JUN-activated kinase; MAPK, mitogen activated protein kinase; MAPKK, mitogen-activated protein kinase kinase; MAPKKK, mitogen activated protein kinase kinase kinase; NF-κB, nuclear factor of κB; STAT, signal transducer and activator of transcription; TEL, translocation-ETS-leukaemia.

Nuclear factors

A third way to block the upregulation of MMPs in human tumours is by targeting the nuclear factors that regulate these genes (Figure 4). Many different extracellular stimuli and signalling pathways that activate MMP expression converge at the AP1 DNA-binding site. Glucocorticoids interact with the AP1 binding site and prevent the upregulation of MMPs [87], but this treatment can affect the expression of many genes and have several side effects. Natural products such as nobiletin — a flavonoid obtained from *Citrus depressa*

— have been shown to inhibit AP1 binding activity and suppress both the production of MMP-1 and MMP-9 by human fibrosarcoma cells and the invasive properties of these cells [88]. Similarly, curcuminoids, natural products of the Indian spice turmeric, inhibit *MMP9* expression by interfering with AP1-induced transcription [89].

Another factor that can be targeted to prevent MMP transcription is NF-κB in cancer [90,91]. Interestingly, PS-341 — a proteasome inhibitor that blocks the degradation of inhibitor of κB (IκB) and thereby maintains NFκB in an inactive status, might be effective in treating multiple myeloma and other cancer types in humans [92]. Synthetic triterpenoids and non-steroidal anti-inflammatory drugs also interfere with the NF-κB pathway [93,94].

Some transcription factors such as p53, PTEN (phosphatase and tensin homologue) and ETS transcription factor (TEL), are involved in negative regulation of MMP expression. Their activity is commonly lost during tumour progression which leads to an increase in the proteolytic capacity of tumour cells [95,96]. Adenoviral delivery of wild-type p53 into squamous-cell carcinoma cells that carry mutant forms of p53 inhibits expression of MMPs and invasive properties, independently of the pro-apoptotic effect of p53 on these cells [97].

Strategies for Blocking proMMP Activation

Inhibitors of plasmin can prevent cleavage of proMMP, and combined with administration of MMPIs, can profoundly reduce tissue, highlighting the potential for similar combinatorial treatment of cancer [98]. MT-MMP, an MMP-activating enzyme, is also overexpressed by different malignant tumours. Together with their general proteolytic behaviour, the MT-MMPs should be considered primary targets.

Anti-inflammatory cytokines, such as IL-4 and IL-13, have been shown to interfere with the proMMP activation process rather than with enzyme expression [99]. Natural products such as green tea catechins, have also been reported to block the MT1-MMP-dependent activation of proMMPs [100]. Similarly, since MMP-3 is a well-characterized activator of proMMPs, inhibitors of this enzyme will also prevent the activation of other proMMPs.

Proprotein convertase inhibitors

A selective furin inhibitor, α1-PDX, has been shown to prevent tumour growth and invasion of human cancer cells [101,102]. Activation of MT1-MMP was prevented by Furin inhibitions, resulting in reduced processing of proMMP-2.Similar results were obtained using a synthetic furin inhibitor [103]. However, the effect of these inhibitors on the activation of the secreted convertase sensitive MMPs, such as MMP-11 (stromelysin-3), which is strongly expressed by tumour stroma [104], has not been reported. Convertase inhibitors block MMPs, but in view of the essential roles of convertases in protein processing in many tissues, side effects could limit the dosage that can be administered, and therefore limit efficacy in humans.

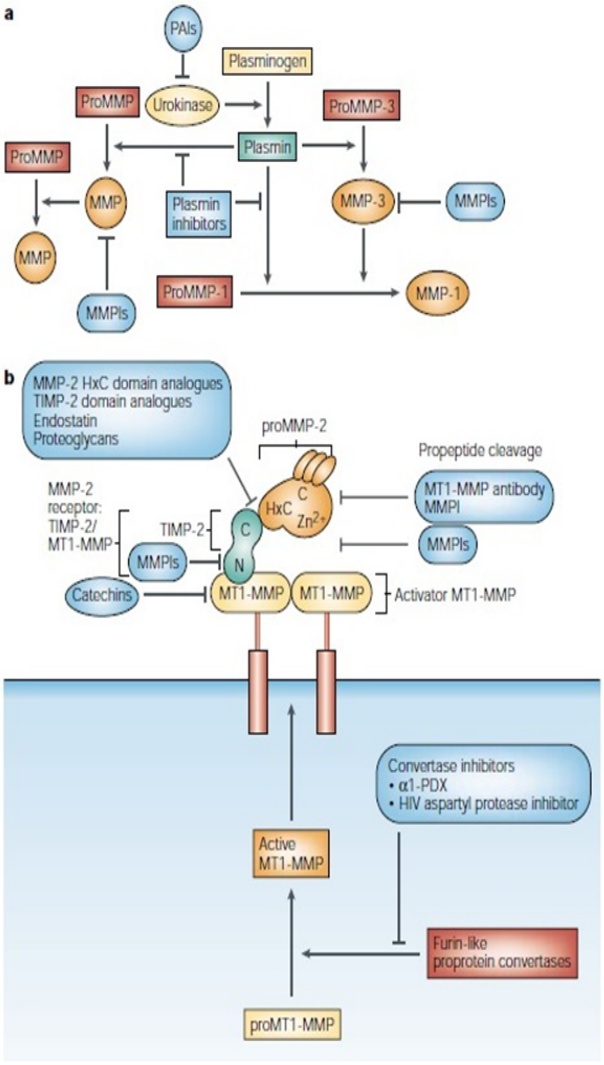


Figure 5: Strategies for blocking proMMP activation.

a | Active MMPs are generated through a multistep proteolytic cascade that involves plasminogen cleavage by urokinase to create plasmin, which then cleaves proMMPs to create active MMPs. Some of these MMPs go on to cleave other proMMPs. MMP inhibitors (MMPIs) block MMP generation from proMMPs. Serine proteinase inhibitors (PAIs) of urokinase block plasmin generation, and plasmin inhibitors block the proMMP conversion to MMPs.

b | Blockade of cell-surface and furin-mediated activation of MMPs. ProMMP-2 activation occurs after the formation of a ternary complex that contains proMMP-2 linked to cell-surface MT1-MMP via a tissue inhibitors of metalloproteinase (TIMP)-2 bridge. The TIMP-2 inhibitory amino (N) domain binds to the active site of MT1-MMP, inhibiting its proteolytic activity. TIMP-2 also binds MMP-2 outside the catalytic domain (C) on the outer rim of the hemopexin carboxy-terminal domain (HxC), at the junction of hemopexin modules III and IV [129]. The crystal structure of the complex proMMP-2/TIMP-2 has shown that the inhibitory N domain of TIMP-2 is not in contact with the HxC domain, and is also distant from the catalytic site of MMP-2 [130]. MT1-MMP has been shown to dimerize in forming the trimolecular complex on the cell surface with MMP-2. MT2-MMP does not require TIMP-2 to bind or activate MMP-2 [108], and it is not known whether the other MT-MMPs form a complex with TIMP-2. Compounds that interfere with the activation of proMMP-2 include catechins, anti-MT1-MMP antibody and MMPIs. MMP-2 HxC domain analogues and TIMP-2 domain analogues disrupt the TIMP-2/MMP2 interaction, and prevent MMP-2 activation. Endostatin and proteoglycans, which form complexes with MMP-2, inhibit processing and activation by MT1-MMP. Active MT1-MMP is generated through a proteolytic cascade from proMT1-MMP, which are generated by furin-like proprotein convertases. These convertases can be inhibited by inhibitors such as α1-PDX and human immunodeficiency virus (HIV) aspartyl protease inhibitors.

Targeting MMP-2 Activation

MMP activation can also be blocked by the use of thrombospondin- 1 — an anti-angiogenic factor that inhibits proMMP-2 and proMMP-9 activation [28,29]. Similarly, thrombospondin-2 promotes MMP-2 endocytosis via the low-density lipoprotein-receptor-related protein pathway [27]. Endostatin forms a complex with MMP-2 and inhibits processing and activation by MT1-MMP [105], which might partially explain its antiangiogenic activity. Similarly proteoglycans, such as testican-3 and its splice-variant gene product N-Tes, can suppress proMMP-2 activation that is mediated by MT-MMPs, with the subsequent abrogation of invasive properties of glioma cells [106]. A potent inhibitor of proMMP-2 activation *in vitro* is recombinant hemopexin, indicating the feasibility of targeting this interaction to block MMP-2 activation *in vivo* [24,107,108]. Analogues of the TIMP-2 C-terminal domain might be used to compete for TIMP-2 binding to the MMP-2 hemopexin C-terminal domain, and so might prevent trimolecular complex formation that is required for MMP-2 activation. Accordingly, MMPIs have both a direct effect in inhibiting active site MMPs and an indirect effect in blocking TIMP-2 binding and MMP-2 activation by MT1-MMP.

Inhibitor	Structure	Specificity	Comments
Marimastat (BB-2516)	Peptido mimetic	Broad spectrum	Survival benefit in a subset of gastric cancer patients Survival benefit in glioblastoma multiforme patients in combination with temozolomide Survival rate similar to gemcitabine in pancreatic cancer No survival benefit in SCL, NSCL and ovarian cancer patients
Tanomastat (BAY 12-9566)	Non-peptido mimetic	MMP-2, 3, 9	Development halted because treated patients showed poorer survival than controls
Prinomastat (AG3340)	Non-peptido mimetic	Broad spectrum	No survival benefits in NSCL cancer patients No difference in progression of prostate carcinomas.
Metastat (COL-3)	Tetracycline derivative	Gelatinases	Multiple mechanisms of action against MMPs Currently recruiting Kaposi's sarcoma patients
Neovastat (AE-941)	Shark cartilage extract	Broad spectrum	Multiple mechanisms of action on MMPs Currently recruiting renal-cell carcinoma, multiple myeloma and NSCL cancer patients
BMS-275291	Non-peptido mimetic	Broad spectrum	Currently recruiting NSCL cancer patients
MMI270	Non-peptido mimetic	Broad spectrum	Anti-angiogenic and anti-metastatic effects in animal models Phase I studies in patients with advanced malignancies

Table 2: Matrix metalloproteinase inhibitors in clinical development for cancer therapy.

MMPs, matrix metalloproteinases; NSCL cancer, non-small-cell lung cancer; SCL cancer, small-cell lung cancer. Table adapted from Coussens et al., 2002; Hidalgo et al., 2001 [64,109].

Natural Compounds as MMP-Inhibitors

Matrix metalloproteinases have been heralded as promising targets for cancer therapy on the basis of their massive up-regulation in malignant tissues and their unique ability to degrade all components of the extracellular matrix. Preclinical studies testing the efficacy of MMP suppression in tumor models were so compelling that synthetic metalloproteinase inhibitors (MPIs) were rapidly developed and routed into human clinical trials. The results of these trials were, however, disappointing.

Natural inhibitors of MMPs

Natural inhibitors of MMPs- TIMPs, were also used to block MMPs activity. Although they have demonstrated efficacy in experimental models, TIMPs may exert MMP-independent promoting effects [3]. To avoid the negative results and toxicity issues raised by the use of synthetic MMPIs, one answer was provided from the field of natural compounds. One compound taken into consideration was extracted from shark cartilage. Oral administration of a standardized extract, neovastat, exerts anti-angiogenic and anti-metastatic activities and these effects depend not only on the inhibition of MMPs enzymatic activity, but also on the inhibition of VEGF. Another natural agent that has anticancer effects is genistein, a soy isoflavonoid, structurally similar to estradiol. Apart from its estrogening and anti-estrogenic properties, genistein confers tumor inhibition growth and invasion effects, interfering with the expression ratio and activity of several MMPs and TIMPs [110].

In 2004, Lambert et. al. reported that the matrix metalloproteinase inhibitors (MMPIs) may be derived from natural resources such as herbs, plants, fruits, and other agriculture products [111]. New and potentially beneficial compounds isolated from these sources were shown to exhibit some degree of MMPI activities, but they were far less potent and specific than the TIMP family. These natural compounds included long chain fatty acids, epigallocatechin gallate (EGCG) and other polyphenols and flavonoids.

Perhaps the most thoroughly studied class of natural MMP inhibitors are the endogenous tissue inhibitors of metalloproteinases (TIMPs), of which four are currently known, designated as TIMP-1 through -4. It is assumed that the natural ratio of MMPs to TIMPs is tightly regulated, and a disruption in the natural balance between these two families is often associated with the progression of multiple disease states. Each of the four TIMPs forms a complex with the MMPs in a 1:1 stoichiometry, exhibiting high affinity, but varying degrees of selectivity.

MMPIs from marine natural products

Marine saccharoid MMPIs: Most of the marine saccharoid MMPIs inhibit MMP by direct down regulation of MMP-9 transcription or via inhibition of activator protein-1(AP-1) pathway or nuclear factor κB (NF-κB) pathway.

Marine saccharoid MMPIs exhibit high MMPs inhibitory activity either by direct inhibition of the enzyme or by inhibiting the expression of MMPs. These have also shown low toxicity levels. However, due to high molecular weight of theses MMPIs, the structure-activity relationship and the mechanism of the activity is hard to be addressed. If these shortcomings are overcome in the future, marine saccharoid MMPIs have a great potential to be used in clinical applications.

Marine flavonoids and polyphenols MMPIs: Flavonoid glycosides, isorhamnetin 3-O-b-D-glucosides, and quercetin 3-O-b-D-glucoside were isolated *from Salicornia herbacea* and their inhibitory effects on matrix metalloproteinase-9 and -2 were evaluated in human fibrosarcoma cell line [112].

Flavonoids and polyphenols MMPIs have excellent MMPs inhibitory activities; however they show a high toxicity level. Therefore, the pharmaceutical applications of these MMPIs are limited. Researchers should pay attention to reduce their toxicity levels by altering the structure in a way that it preserves the bioactivity. Then this class of MMPIs will gain a huge potential to be used in clinical applications.

Marine fatty acid MMPIs: Long-chain fatty acids can inhibit MMPs. However for different MMPs, the degree of inhibition is different. Oleic acid and elaidic acid can inhibit MMP-2 and MMP-9 with the micromole Ki values, although their inhibitory effects on collagenase-1 (MMP-1) are weak [113]. The fatty acid chain length and its degree of saturation is related to the level of inhibition, as the fatty acids with long carbon chains show stronger inhibition than the short ones, and the non-saturation degree shows a positive correlation to the overall inhibitory capacity of the fatty acid chains. Fatty acids bind to neutrophil elastase, while the parinaric acids are inhibitors of neutrophil elastase. The fatty acids bind to plasmin for example, oleic acid can modulate fibrinolysis. It is well known that the marine fishes are rich in omega-3 long-chain polyunsaturated fatty acids (ω3 LC-PUFAs), especially eicosapentaenoic (EPA) and docosahexaenoic acid (DHA). Suzuki et al. found that the inhibition of lung metastasis of a colon cancer cell line by EPA and DHA was associated with a reduced activity of MMP-9. A wide range of biological activities such as cytotoxicity [114], antimicrobial [115], antifouling [116], and enzyme inhibition are shown by acetylenic fatty acids isolated from marine sponges. Sodium 1-(12-hydroxy) octadecanyl sulfate inhibits MMP-2. Callysponginol sulfate A, extracted from the marine sponge, *Callyspongia truncate*, inhibits recombinant MT1-MMP.

Other marine natural products MMPIs: The compounds extracted from shark cartilage (such as NeovastatÒ, AE-941, U-995) have been investigated for their potential use as MMPIs. NeovastatÒ inhibits enzymatic activity of MMP-2 with minor inhibition of MMP-1, -7, -9 and -13. TIMP-like proteins within AE-941 could be responsible for its specific MMP inhibitory property [117]. The Atlantic cod (Gadus morhua) muscle contains a 21-kDa proteinase inhibitor which has properties similar to human TIMP-2. The inhibitor was found to inhibit the gelatin-degrading enzymes present in the gelatin-bound fraction. In addition, it inhibited gelatinolytic activity obtained from a human macrophage cell medium rich in MMP-9 [118].

Marine plants: Metabolites from marine plants have outstanding biological activities. A highly effective anticoagulant and antiproliferative agent, sulfated polysaccharide, from brown alga *Ecklonia cava*, exhibited a promising antiproliferative effect on human promyelocytic leukemia (HL-60) and human leukemic monocyte lymphoma (U-937) cells. Fucoidan extracts from sea weed *Cladosiphon novae-* reduce the cellular invasiveness in human fibrosarcoma HT1080 cells by suppressing the activity of MMP-2 and MMP-9. Further, it has been reported that these fucoidan extracts suppresse the expression and secretion of an angiogenesis factor, vascular endothelial growth factor (VEGF); thereby reporting the inhibitory effects on invasion and angiogenesis of tumor cells [119].

Extracts from *Eisenia bicyclis*, *Ecklonia cava*, and *Ecklonia stolonifera* have strongly reduced MMP-1 expression via inhibiting both NF-kappa B and AP-1 reporter. Free radical scavenging activity of phlorotannins from *Ecklonia* species has been reported. Dieckol from marine brown alga, *E. cava* has been reported to suppress LPS-induced production of nitric oxide (NO), prostaglandin E2 (PGE2), inducible nitric oxide synthase (iNOS) and cyclooxygenase-2 (COX-2) in murine BV2 microglia; thus establishing dieckol as a potent anti-inflammatory and neuroprotective agent. Kong et al. reported that flavonoid glycosides, isolated from this plant have inhibited the expression of MMP-2 and MMP-9 and elevated the TIMP-1 expression in human fibrosarcoma (HT1080) cells. Moreover, the down regulation of MMP-9 and MMP-2 by these flavonoids is due to the interference with the transcription factor AP-1, there by suggesting that these flavonoid glycosides can be used as potent natural chemopreventive agents for cancer. Phlorotannins from brown alga *E. cava* have been reported to have inhibitory activity on MMP-2 and MMP-9, signifying the role of phlorotannins as potential and safe marine derived MMPIs.

Miscellaneous natural products

Screening has led to the discovery of both synthetic and natural MMP inhibitors. The latter include tetracyclines, such as doxycycline and minocycline, for which it has been found that chemical modification can separate MMP activity from antibiotic activity [120]. Actinonin has been identified as an MMP inhibitor and it is a succinyl hydroxamic acid that bears close structural similarity to compounds obtained by substrate based design. Although significant advances have been made in inhibitor design, it is still not clear how to design compounds that specifically inhibit individual MMPs inspite of the available structural data. This remains a major challenge for MMP inhibitor medicinal chemistry [121].

Other natural compounds

Fujita et al., 2003 reported that ageladine A, a fluorescent alkaloid isolated from the marine sponge *Agelas nakamurai*, inhibits MMP-1, -8, -9, -12, and -13.This compound could also inhibit MMP-2but N-methylated derivatives did not inhibit MMP-2. The inhibition is not due to Zn^{2+} chelation, as ageladine is not capable of chelating to Zn^{2+}, and a kinetic analysis indicated that the inhibition was not competitive. In addition, bovine aortic endothelial cell migration and vascular formation by murine ES cells were significantly inhibited by this compound [122].

Future Perspective

Research into neoepitopes will provide important and novel means of diagnosis, prognosis and increasing treatment efficacy in cancer. However, to fully take advantage of neoepitopes as highly valuable cancer biomarkers, it is very important to understand the physiological mechanisms and signalling pathways that regulate their generation. Thus, the ultimate goal of new diagnostic tests should be to use highly reliable non-invasive mechanism-based biomarkers. At present, receptors, cell adhesion molecules, growth factors and enzymes, with their related protein substrates (e.g., MMPs and extracellular matrix components), are all hot research areas in the development of cancer drugs and diagnostic assays [123].

In the past year, further evidence establishing the usefulness of β interferons and glatiramer in the treatment of relapsing-remitting multiple sclerosis has been advanced. Interferon-β-1b was also shown to be efficacious in secondary progressive multiple sclerosis. There are more than 20 MMPs identified that share several common features: signal peptides, prodomain, and prodomain, and catalytic domain, with at least eight of these proteins clustered on chromosome 11 (MMPs -1, -3, -7, -8, -10, -12, -13, and -20), probably due to a gene duplication event [124]. Johnson, PR et al., reported in 2001, that although the healthy adult lung is not a major source of MMPs, parenchymal cells such as airway epithelium, fibroblast, and smooth muscle have the capacity to express active MMPs following stimulation by a variety of agents such as infectious pathogens, environmental toxins, growth factors, and cytokines [125]. Lopez-Boado et al. 2000 reported a 25-fold induction of MMP-7 in the lung epithelial cells following infection with *Escherichia coli* and *Pseudomonas aeruginosa*, which could explain the up regulation of this enzyme in the airway of cystic fibrosis patients who are commonly infected with these bacteria. It also has been shown that proinflammatory cytokines such as interleukin 1 beta (IL-1β) and tumor necrosis factor alpha (TNF-α), upregulate the expression of MMP-9 in human airway epithelial cells following a 1-day treatment [126]. Additionally, inflammatory cells invading the lung during the course of COPD are also a major source of different MMPs. It has been shown that the neutrophils and macrophages, the predominant inflammatory cells in the lungs of COPD patients, have the capacity to release MMPs -2, -3, -7, -9, and -12 [127].

Future of Cell Utilization for Disc Disease

Despite the growing number of research data on cell-based experimental therapy for IVD disease, it is clear that we do not know much about native disc cells and their microenvironment. This lack of information is a major obstacle in building a strategy for the treatment of IVD disease. Investigating the specific differentiation status of native IVD cells and their homeostasis will surely provide more ideas and clues for efficient therapeutic approaches. Although cell-based therapy for IVD disease is still in its infancy, the stage of testing a variety of cells for injection should be toned. To progress to the next step, we should be investigating what exactly IVD cells are, and how they control their homeostasis, along with various studies optimising parameters, such as cell dosage and culture period and the severity of IVD degeneration in the recipient [128].

Moreover, there is the need for attention to the stage and type of cancer that is likely to be evaluated in clinical versus preclinical studies. For example, the selection of advanced pancreatic and lung cancers for clinical trials was based on considerations such as expected survival time and patient availability, both of which affect the time and financial resources required to achieve statistically significant results. Patent issues, competition, and impatience contributed to the decision to proceed at an unprecedented pace in an inappropriate setting, and these factors will undoubtedly continue to influence drug development decisions in the future.

References

1. Van Lint P, Libert C (2007) Chemokine and cytokine processing by matrix metalloproteinases and its effect on leukocyte migration and inflammation. J Leukoc Biol 82: 1375-1381.
2. Gross J, Lapiere C (1962) Collagenolytic activity in amphibian tissues: A tissue culture assay. Proc Natl Acad Sci USA 48: 1014–1022.
3. Brew K, Nagase H (2010) The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity. Biochim Biophys Acta 1803: 55-71.
4. Cawston TE (1996) Metalloproteinase inhibitors and the prevention of connective tissue breakdown. Pharmacol Ther 70: 163-182.
5. Johnson LL, Dyer R, Hupe DJ (1998) Matrix metalloproteinases. Curr Opin Chem Biol 2: 466-471.
6. Massova I, Kotra LP, Fridman R, Mobashery S (1998) Matrix metalloproteinases: structures, evolution, and diversification. FASEB J 12: 1075-1095.
7. Sternlicht MD, Werb Z (2001) How matrix metalloproteinases regulate cell behavior. Annu Rev Cell Dev Biol 17: 463-516.
8. Edwards DR (2001) In: Clendeninn NJ & Appelt K (eds.). Matrix Metalloproteinase Inhibitors in Cancer Therapy. Humana Press, Totowa, New Jersey, USA.
9. Murphy G (2008) The ADAMs: signalling scissors in the tumour microenvironment. Nat Rev Cancer 8: 929-941.
10. Brooks PC, Strömblad S, Sanders LC, von Schalscha TL, Aimes RT, et al. (1996) Localization of matrix metalloproteinase MMP-2 to the surface of invasive cells by interaction with integrin alpha v beta 3. Cell 85: 683-693.
11. Yu Q, Stamenkovic I (1999) Localization of matrix metalloproteinase 9 to the cell surface provides a mechanism for CD44-mediated tumor invasion. Genes Dev 13: 35-48.
12. Yu WH, Woessner JF Jr, McNeish JD, Stamenkovic I (2002) CD44 anchors the assembly of matrilysin/MMP-7 with heparin-binding epidermal growth factor precursor and ErbB4 and regulates female reproductive organ remodeling. Genes Dev 16: 307-323.
13. Bischof P, Meisser A, Campana A (2002) Control of MMP-9 expression at the maternal-fetal interface. J Reprod Immunol 55: 3-10.
14. Muller D, Quantin B, Gesnel MC, Millon-Collard R, Abecassis J, et al. (1988) The collagenase gene family in humans consists of at least four members. Biochem J 253: 187-192.
15. Jung JY, Warter S, Rumpler Y (1990) Localization of stromelysin 2 gene to the q22.3-23 region of chromosome 11 by *in situ* hybridization. Ann Genet 33: 21-23.
16. Levy A, Zucman J, Delattre O, Mattei MG, Rio MC, et al. (1992) Assignment of the human stromelysin 3 (STMY3) gene to the q11.2 region of chromosome 22. Genomics 13: 881-883.
17. Anglard P, Melot T, Guérin E, Thomas G, Basset P (1995) Structure and promoter characterization of the human stromelysin-3 gene. J Biol Chem 270: 20337-20344.
18. Luo D, Mari B, Stoll I, Anglard P (2002) Alternative splicing and promoter usage generates an intracellular stromelysin 3 isoform directly translated as an active matrix metalloproteinase. J Biol Chem 277: 25527-25536.

19. Meyer BS, Rademann J (2012) Extra- and intracellular imaging of human matrix metalloprotease 11 (hMMP-11) with a cell-penetrating FRET substrate. J Biol Chem 287: 37857-37867.

20. Freije JM, Díez-Itza I, Balbín M, Sánchez LM, Blasco R, et al. (1994) Molecular cloning and expression of collagenase-3, a novel human matrix metalloproteinase produced by breast carcinomas. J Biol Chem 269: 16766-16773.

21. Kudo Y, Iizuka S, Yoshida M, Tsunematsu T, Kondo T, et al. (2012) Matrix metalloproteinase-13 (MMP-13) directly and indirectly promotes tumor angiogenesis. J Biol Chem 287: 38716-38728.

22. Yu G, Kovkarova-Naumovski E, Jara P, Parwani A, Kass D, et al. (2012) Matrix metalloproteinase-19 is a key regulator of lung fibrosis in mice and humans. Am J Respir Crit Care Med 186: 752-762.

23. Saarialho-Kere U, Kerkelä E, Jahkola T, Suomela S, Keski-Oja J, et al. (2002) Epilysin (MMP-28) expression is associated with cell proliferation during epithelial repair. J Invest Dermatol 119: 14-21.

24. Strongin AY, Collier I, Bannikov G, Marmer BL, Grant GA, et al. (1995) Mechanism of cell surface activation of 72-kDa type IV collagenase. Isolation of the activated form of the membrane metalloprotease. J Biol Chem 270: 5331-5338.

25. Deryugina EI, Ratnikov B, Monosov E, Postnova TI, DiScipio R, et al. (2001) MT1-MMP initiates activation of pro-MMP-2 and integrin alphavbeta3 promotes maturation of MMP-2 in breast carcinoma cells. Exp Cell Res 263: 209-223.

26. Sottrup-Jensen L, Birkedal-Hansen H (1989) Human fibroblast collagenase-alpha-macroglobulin interactions. Localization of cleavage sites in the bait regions of five mammalian alpha-macroglobulins. J Biol Chem 264: 393-401.

27. Yang Z, Strickland DK, Bornstein P (2001) Extracellular matrix metalloproteinase 2 levels are regulated by the low density lipoprotein-related scavenger receptor and thrombospondin 2. J Biol Chem 276: 8403-8408.

28. Bein K, Simons M (2000) Thrombospondin type 1 repeats interact with matrix metalloproteinase 2. Regulation of metalloproteinase activity. J Biol Chem 275: 32167-32173.

29. Rodriguez-Manzaneque JC, Lane TF, Ortega MA, Hynes RO, Lawler J, et al. (2001) Thrombospondin-1 suppresses spontaneous tumor growth and inhibits activation of matrix metalloproteinase-9 and mobilization of vascular endothelial growth factor. Proc Natl Acad Sci U S A 98: 12485-12490.

30. Wang Z, Juttermann R, Soloway PD (2000) TIMP-2 is required for efficient activation of proMMP-2 *in vivo*. J Biol Chem 275: 26411-26415.

31. Streuli C (1999) Extracellular matrix remodelling and cellular differentiation. Curr Opin Cell Biol 11: 634-640.

32. Giannelli G, Falk-Marzillier J, Schiraldi O, Stetler-Stevenson WG, Quaranta V (1997) Induction of cell migration by matrix metalloprotease-2 cleavage of laminin-5. Science 277: 225-228.

33. Xu J, Rodriguez D, Petitclerc E, Kim JJ, Hangai M, et al. (2001) Proteolytic exposure of a cryptic site within collagen type IV is required for angiogenesis and tumor growth *in vivo*. J Cell Biol 154: 1069-1079.

34. Mañes S, Mira E, Barbacid MM, Ciprés A, Fernández-Resa P, et al. (1997) Identification of insulin-like growth factor-binding protein-1 as a potential physiological substrate for human stromelysin-3. J Biol Chem 272: 25706-25712.

35. Mañes S, Llorente M, Lacalle RA, Gómez-Moutón C, Kremer L, et al. (1999) The matrix metalloproteinase-9 regulates the insulin-like growth factor-triggered autocrine response in DU-145 carcinoma cells. J Biol Chem 274: 6935-6945.

36. Whitelock JM, Murdoch AD, Iozzo RV, Underwood PA (1996) The degradation of human endothelial cell-derived perlecan and release of bound basic fibroblast growth factor by stromelysin, collagenase, plasmin, and heparanases. J Biol Chem 271: 10079-10086.

37. Peschon JJ, Slack JL, Reddy P, Stocking KL, Sunnarborg SW, et al. (1998) An essential role for ectodomain shedding in mammalian development. Science 282: 1281-1284.

38. Yu Q, Stamenkovic I (2000) Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-beta and promotes tumor invasion and angiogenesis. Genes Dev 14: 163-176.

39. Levi E, Fridman R, Miao HQ, Ma YS, Yayon A, et al. (1996) Matrix metalloproteinase 2 releases active soluble ectodomain of fibroblast growth factor receptor 1. Proc Natl Acad Sci U S A 93: 7069-7074.

40. Codony-Servat J, Albanell J, Lopez-Talavera JC, Arribas J, Baselga J (1999) Cleavage of the HER2 ectodomain is a pervanadate-activable process that is inhibited by the tissue inhibitor of metalloproteases-1 in breast cancer cells. Cancer Res 59: 1196-1201.

41. Vecchi M, Rudolph-Owen LA, Brown CL, Dempsey PJ, Carpenter G (1998) Tyrosine phosphorylation and proteolysis. Pervanadate-induced, metalloprotease-dependent cleavage of the ErbB-4 receptor and amphiregulin. J Biol Chem 273: 20589-20595.

42. Nath D, Williamson NJ, Jarvis R, Murphy G (2001) Shedding of c-Met is regulated by crosstalk between a G-protein coupled receptor and the EGF receptor and is mediated by a TIMP-3 sensitive metalloproteinase. J Cell Sci 114: 1213-1220.

43. Noë V, Fingleton B, Jacobs K, Crawford HC, Vermeulen S, et al. (2001) Release of an invasion promoter E-cadherin fragment by matrilysin and stromelysin-1. J Cell Sci 114: 111-118.

44. Kajita M, Itoh Y, Chiba T, Mori H, Okada A, et al. (2001) Membrane-type 1 matrix metalloproteinase cleaves CD44 and promotes cell migration. J Cell Biol 153: 893-904.

45. Westermarck J, Kähäri VM (1999) Regulation of matrix metalloproteinase expression in tumor invasion. FASEB J 13: 781-792.

46. Chandler S, Coates R, Gearing A, Lury J, Wells G, et al. (1995) Matrix metalloproteinases degrade myelin basic protein. Neurosci Lett 201: 223-226.

47. George SJ (2000) Therapeutic potential of matrix metalloproteinase inhibitors in atherosclerosis. Expert Opin Investig Drugs 9: 993-1007.

48. Arthur MJ (2000) Fibrogenesis II. Metalloproteinases and their inhibitors in liver fibrosis. Am J Physiol Gastrointest Liver Physiol 279: G245-249.

49. Thompson RW, Parks WC (1996) Role of matrix metalloproteinases in abdominal aortic aneurysms. Ann N Y Acad Sci 800: 157-174.

50. Saarialho-Kere UK, Vaalamo M, Puolakkainen P, Airola K, Parks WC, et al. (1996) Enhanced expression of matrilysin, collagenase, and stromelysin-1 in gastrointestinal ulcers. Am J Pathol 148: 519-526.

51. Rosenberg GA (1995) Matrix metalloproteinases in brain injury. J Neurotrauma 12: 833-842.

52. Golub LM, Wolff M, Roberts S, Lee HM, Leung M, et al. (1994) Treating periodontal diseases by blocking tissue-destructive enzymes. J Am Dent Assoc 125: 163-169.

53. Corbel M, Boichot E, Lagente V (2000) Role of gelatinases MMP-2 and MMP-9 in tissue remodeling following acute lung injury. Braz J Med Biol Res 33: 749-754.

54. Konttinen YT, Ainola M, Valleala H, Ma J, Ida H, et al. (1999) Analysis of 16 different matrix metalloproteinases (MMP-1 to MMP-20) in the synovial membrane: different profiles in trauma and rheumatoid arthritis. Ann Rheum Dis 58: 691-697.

55. Hua J, Muschel RJ (1996) Inhibition of matrix metalloproteinase 9 expression by a ribozyme blocks metastasis in a rat sarcoma model system. Cancer Res 56: 5279-5284.

56. Yonemura Y, Endo Y, Fujita H, Kimura K, Sugiyama K, et al. (2001) Inhibition of peritoneal dissemination in human gastric cancer by MMP-7-specific antisense oligonucleotide. J Exp Clin Cancer Res 20: 205-212.

57. Kondraganti S, Mohanam S, Chintala SK, Kin Y, Jasti SL, et al. (2000) Selective suppression of matrix metalloproteinase-9 in human glioblastoma cells by antisense gene transfer impairs glioblastoma cell invasion. Cancer Res 60: 6851-6855.

58. Noonberg SB, Benz CC (2000) Tyrosine kinase inhibitors targeted to the epidermal growth factor receptor subfamily: role as anticancer agents. Drugs 59: 753-767.

59. Elkin M, Reich R, Nagler A, Aingorn E, Pines M, et al. (1999) Inhibition of matrix metalloproteinase-2 expression and bladder carcinoma metastasis by halofuginone. Clin Cancer Res 5: 1982-1988.

60. Silletti S, Kessler T, Goldberg J, Boger DL, Cheresh DA (2001) Disruption of matrix metalloproteinase 2 binding to integrin alpha vbeta 3 by an organic molecule inhibits angiogenesis and tumor growth *in vivo*. Proc Natl Acad Sci U S A 98: 119-124.

61. Liu S, Netzel-Arnett S, Birkedal-Hansen H, Leppla SH (2000) Tumor cell-selective cytotoxicity of matrix metalloproteinase-activated anthrax toxin. Cancer Res 60: 6061-6067.

62. Jiang Y, Wang M, Celiker MY, Liu YE, Sang QX, et al. (2001) Stimulation of mammary tumorigenesis by systemic tissue inhibitor of matrix metalloproteinase 4 gene delivery. Cancer Res 61: 2365-2370.

63. Moore MJ et al. (2002) A comparison between gemcitabine (GEM) and the matrix metalloproteinase (MMP) inhibitor BAY12–9566 (9566) in patients (PTS) with advanced pancreatic cancer. ASCO Online.

64. Hidalgo M, Eckhardt SG (2001) Development of matrix metalloproteinase inhibitors in cancer therapy. J Natl Cancer Inst 93: 178-193.

65. Koivunen E, Arap W, Valtanen H, Rainisalo A, Medina OP, et al. (1999) Tumor targeting with a selective gelatinase inhibitor. Nat Biotechnol 17: 768-774.

66. Boissier S, Ferreras M, Peyruchaud O, Magnetto S, Ebetino FH, et al. (2000) Bisphosphonates inhibit breast and prostate carcinoma cell invasion, an early event in the formation of bone metastases. Cancer Res 60: 2949-2954.

67. Falardeau P, Champagne P, Poyet P, Hariton C, Dupont E (2001) Neovastat, a naturally occurring multifunctional antiangiogenic drug, in phase III clinical trials. Semin Oncol 28: 620-625.

68. Garbisa S, Biggin S, Cavallarin N, Sartor L, Benelli R, et al. (1999) Tumor invasion: molecular shears blunted by green tea. Nat Med 5: 1216.

69. Jiang MC, Liao CF, Lee PH (2001) Aspirin inhibits matrix metalloproteinase-2 activity, increases E-cadherin production, and inhibits *in vitro* invasion of tumor cells. Biochem Biophys Res Commun 282: 671-677.

70. Nagavarapu U, Relloma K, Herron GS (2002) Membrane type 1 matrix metalloproteinase regulates cellular invasiveness and survival in cutaneous epidermal cells. J Invest Dermatol 118: 573-581.

71. Ala-aho R, Johansson N, Grénman R, Fusenig NE, López-Otín C, et al. (2000) Inhibition of collagenase-3 (MMP-13) expression in transformed human keratinocytes by interferon-gamma is associated with activation of extracellular signal-regulated kinase-1,2 and STAT1. Oncogene 19: 248-257.

72. Ma Z, Qin H, Benveniste EN (2001) Transcriptional suppression of matrix metalloproteinase-9 gene expression by IFN-gamma and IFN-beta: critical role of STAT-1alpha. J Immunol 167: 5150-5159.

73. Slaton JW, Karashima T, Perrotte P, Inoue K, Kim SJ, et al. (2001) Treatment with low-dose interferon-alpha restores the balance between matrix metalloproteinase-9 and E-cadherin expression in human transitional cell carcinoma of the bladder. Clin Cancer Res 7: 2840-2853.

74. Mengshol JA, Mix KS, Brinckerhoff CE (2002) Matrix metalloproteinases as therapeutic targets in arthritic diseases: bull's-eye or missing the mark? Arthritis Rheum 46: 13-20.

75. Lal A, Glazer CA, Martinson HM, Friedman HS, Archer GE, et al. (2002) Mutant epidermal growth factor receptor up-regulates molecular effectors of tumor invasion. Cancer Res 62: 3335-3339.

76. Overall CM, Wrana JL, Sodek J (1989) Independent regulation of collagenase, 72-kDa progelatinase, and metalloendoproteinase inhibitor expression in human fibroblasts by transforming growth factor-beta. J Biol Chem 264: 1860-1869.

77. Uría JA, Jiménez MG, Balbín M, Freije JM, López-Otín C (1998) Differential effects of transforming growth factor-beta on the expression of collagenase-1 and collagenase-3 in human fibroblasts. J Biol Chem 273: 9769-9777.

78. Overall CM (1995) Repression of tissue inhibitor of matrix metalloproteinase expression by all-trans-retinoic acid in rat bone cell populations: comparison with transforming growth factor-beta 1. J Cell Physiol 164: 17-25.

79. Maria JG Jiménez, Milagros Balbín, Jesús Alvarez, Toshihisa Komori, Paolo Bianco, et al. (2001) A regulatory cascade involving retinoic acid, Cbfa1, and matrix metalloproteinases is coupled to the development of a process of perichondrial invasion and osteogenic differentiation during bone formation. J Cell Biol 155: 1333–1344.

80. Muraoka RS, Dumont N, Ritter CA, Dugger TC, Brantley DM, et al. (2002) Blockade of TGF-beta inhibits mammary tumor cell viability, migration, and metastases. J Clin Invest 109: 1551-1559.

81. McGaha TL, Phelps RG, Spiera H, Bona C (2002) Halofuginone, an inhibitor of type-I collagen synthesis and skin sclerosis, blocks transforming-growth-factor-beta-mediated Smad3 activation in fibroblasts. J Invest Dermatol 118: 461-470.

82. Simon C, Goepfert H, Boyd D (1998) Inhibition of the p38 mitogen-activated protein kinase by SB 203580 blocks PMA-induced Mr 92,000 type IV collagenase secretion and *in vitro* invasion. Cancer Res 58: 1135-1139.

83. Johansson N, Ala-aho R, Uitto V, Grénman R, Fusenig NE, et al. (2000) Expression of collagenase-3 (MMP-13) and collagenase-1 (MMP-1) by transformed keratinocytes is dependent on the activity of p38 mitogen-activated protein kinase. J Cell Sci 113: 227-235.

84. Shin M, Yan C, Boyd D (2002) An inhibitor of c-jun aminoterminal kinase (SP600125) represses c-Jun activation, DNA-binding and PMA-inducible 92-kDa type IV collagenase expression. Biochim Biophys Acta 1589: 311-316.

85. Futamura M, Kamiya S, Tsukamoto M, Hirano A, Monden Y, et al. (2001) Malolactomycin D, a potent inhibitor of transcription controlled by the Ras responsive element, inhibits Ras-mediated transformation activity with suppression of MMP-1 and MMP-9 in NIH3T3 cells. Oncogene 20: 6724-6730.

86. Zhang Y, Thant AA, Machida K, Ichigotani Y, Naito Y, et al. (2002) Hyaluronan-CD44s signaling regulates matrix metalloproteinase-2 secretion in a human lung carcinoma cell line QG90. Cancer Res 62: 3962-3965.

87. Karin M, Chang L (2001) AP-1--glucocorticoid receptor crosstalk taken to a higher level. J Endocrinol 169: 447-451.

88. Sato T et al. (2002) Inhibition of activator protein-1 binding activity and phosphatidylinositol 3-kinase pathway by nobiletin, a polymethoxy flavonoid, results in augmentation of tissue inhibitor of metalloproteinases-1 production and suppression of production of matrix metalloproteinases-1 and -9 in human fibrosarcoma HT-1080 cells. Cancer Res 62: 1025–1029.

89. Mohan R, Sivak J, Ashton P, Russo LA, Pham BQ, et al. (2000) Curcuminoids inhibit the angiogenic response stimulated by fibroblast growth factor-2, including expression of matrix metalloproteinase gelatinase B. J Biol Chem 275: 10405-10412.

90. Bond M, Baker AH, Newby AC (1999) Nuclear factor kappaB activity is essential for matrix metalloproteinase-1 and -3 upregulation in rabbit dermal fibroblasts. Biochem Biophys Res Commun 264: 561-567.

91. Han YP, Tuan TL, Wu H, Hughes M, Garner WL (2001) TNF-alpha stimulates activation of pro-MMP2 in human skin through NF-(kappa)B mediated induction of MT1-MMP. J Cell Sci 114: 131-139.

92. Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, et al. (1999) Proteasome inhibitors: a novel class of potent and effective antitumor agents. Cancer Res 59: 2615-2622.

93. Mix KS, Mengshol JA, Benbow U, Vincenti MP, Sporn MB, et al. (2001) A synthetic triterpenoid selectively inhibits the induction of matrix metalloproteinases 1 and 13 by inflammatory cytokines. Arthritis Rheum 44: 1096-1104.

94. Pan MR, Chuang LY, Hung WC (2001) Non-steroidal anti-inflammatory drugs inhibit matrix metalloproteinase-2 expression via repression of transcription in lung cancer cells. FEBS Lett 508: 365-368.

95. Sun Y, Sun Y, Wenger L, Rutter JL, Brinckerhoff CE, et al. (1999) p53 down-regulates human matrix metalloproteinase-1 (Collagenase-1) gene expression. J Biol Chem 274: 11535-11540.

96. Sun Y, Cheung JM, Martel-Pelletier J, Pelletier JP, Wenger L, et al. (2000) Wild type and mutant p53 differentially regulate the gene expression of human collagenase-3 (hMMP-13). J Biol Chem 275: 11327-11332.

97. Ala-aho R, Grénman R, Seth P, Kähäri VM (2002) Adenoviral delivery of p53 gene suppresses expression of collagenase-3 (MMP-13) in squamous carcinoma cells. Oncogene 21: 1187-1195.

98. Lund LR, Romer J, Bugge TH, Nielsen BS, Frandsen TL, et al. (1999) Functional overlap between two classes of matrix-degrading proteases in wound healing. EMBO J 18: 4645-4656.

99. Van Lent PL, Holthuysen AE, Sloetjes A, Lubberts E, Van den Berg WB (2002) Local overexpression of adeno-viral IL-4 protects cartilage from metallo proteinase-induced destruction during immune complex-mediated arthritis by preventing activation of pro-MMPs. Osteoarthritis Cartilage 10: 234-243.

100. Annabi B, Lachambre MP, Bousquet-Gagnon N, Page M, Gingras D, et al. (2002) Green tea polyphenol (-)-epigallocatechin 3-gallate inhibits MMP-2 secretion and MT1-MMP-driven migration in glioblastoma cells. Biochim Biophys Acta 1542: 209-220.

101. Bassi DE, Lopez De Cicco R, Mahloogi H, Zucker S, Thomas G, et al. (2001) Furin inhibition results in absent or decreased invasiveness and tumorigenicity of human cancer cells. Proc Natl Acad Sci U S A 98: 10326-10331.

102. Khatib AM, Siegfried G, Chrétien M, Metrakos P, Seidah NG (2002) Proprotein convertases in tumor progression and malignancy: novel targets in cancer therapy. Am J Pathol 160: 1921-1935.

103. Maquoi E, Noël A, Frankenne F, Angliker H, Murphy G, et al. (1998) Inhibition of matrix metalloproteinase 2 maturation and HT1080 invasiveness by a synthetic furin inhibitor. FEBS Lett 424: 262-266.

104. Basset P, Bellocq JP, Wolf C, Stoll I, Hutin P, et al. (1990) A novel metalloproteinase gene specifically expressed in stromal cells of breast carcinomas. Nature 348: 699-704.

105. Kim YM, Jang JW, Lee OH, Yeon J, Choi EY, et al. (2000) Endostatin inhibits endothelial and tumor cellular invasion by blocking the activation and catalytic activity of matrix metalloproteinase. Cancer Res 60: 5410-5413.

106. Nakada M, Yamada A, Takino T, Miyamori H, Takahashi T, et al. (2001) Suppression of membrane-type 1 matrix metalloproteinase (MMP)-mediated MMP-2 activation and tumor invasion by testican 3 and its splicing variant gene product, N-Tes. Cancer Res 61: 8896-8902.

107. Overall CM, Tam E, McQuibban GA, Morrison C, Wallon UM, et al. (2000) Domain interactions in the gelatinase A.TIMP-2.MT1-MMP activation complex. The ectodomain of the 44-kDa form of membrane type-1 matrix metalloproteinase does not modulate gelatinase A activation. J Biol Chem 275: 39497-39506.

108. Morrison CJ, Butler GS, Bigg HF, Roberts CR, Soloway PD, et al. (2001) Cellular activation of MMP-2 (gelatinase A) by MT2-MMP occurs via a TIMP-2-independent pathway. J Biol Chem 276: 47402-47410.

109. Coussens LM, Fingleton B, Matrisian LM (2002) Matrix metalloproteinase inhibitors and cancer: trials and tribulations. Science 295: 2387-2392.

110. Huang X, Chen S, Xu L, Liu Y, Deb DK, et al. (2005) Genistein inhibits p38 map kinase activation, matrix metalloproteinase type 2, and cell invasion in human prostate epithelial cells. Cancer Res 65: 3470-3478.

111. Lambert E, Dassé E, Haye B, Petitfrère E (2004) TIMPs as multifacial proteins. Crit Rev Oncol Hematol 49: 187-198.

112. Wang S, Cheng Y, Wang F, Sun L, Liu C, et al. (2008) Inhibition activity of sulfated polysaccharide of Sepiella maindroni ink on matrix metalloproteinase (MMP)-2. Biomed Pharmacother 62: 297-302.

113. Berton A, Rigot V, Huet E, Decarme M, Eeckhout Y, et al. (2001) Involvement of fibronectin type II repeats in the efficient inhibition of gelatinases A and B by long-chain unsaturated fatty acids. J Biol Chem 276: 20458-20465.

114. Dai JR, Hallock YF, Cardellina JH II, Gray GN, Boyd MR (1996) Triangulynes A-H and trangulynic acid, new cytotxic polyacetylenes from the marine sponge Pellina triangulata. J Nat Prod 59: 860-865.

115. Li HY, Matsunaga S, Fusetani N (1994) Corticatic acids A-C, antifungal acetylenic acids from the marine sponge, Petrosia corticata. J Nat Prod 57: 1464-1467.

116. Tsukamoto S, Kato H, Hirota H, Fusetani N (1997) Mauritiamine, a new antifouling oroidin dimer from the marine sponge Agelas mauritiana. J Nat Prod 60: 126–130.

117. Gingras D, Renaud A, Mousseau N, Beaulieu E, Kachra Z, et al. (2001) Matrix proteinase inhibition by AE-941, a multifunctional antiangiogenic compound. Anticancer Res 21: 145-155.

118. Lødemel JB, Egge-Jacobsen W, Olsen RL (2004) Detection of TIMP-2-like protein in Atlantic cod (Gadus morhua) muscle using two-dimensional real-time reverse zymography. Comp Biochem Physiol B Biochem Mol Biol 139: 253-259.

119. Huxley-Jones J, Clarke TK, Beck C, Toubaris G, Robertson DL, et al. (2007) The evolution of the vertebrate metzincins; insights from Ciona intestinalis and Danio rerio. BMC Evol Biol 7: 63.

120. Suomalainen K, Sorsa T, Golub LM, Ramamurthy N, Lee HM, et al. (1992) Specificity of the anticollagenase action of tetracyclines: relevance to their anti-inflammatory potential. Antimicrob Agents Chemother 36: 227-229.

121. Lelièvre Y, Bouboutou R, Boiziau J, Cartwright T (1989) [Inhibition of synovial collagenase by actinonin. Study of structure/activity relationship]. Pathol Biol (Paris) 37: 43-46.

122. Fujita M, Nakao Y, Matsunaga S, Seiki M, Itoh Y, et al. (2003) Ageladine A: an antiangiogenic matrixmetalloproteinase inhibitor from the marine sponge *Agelas nakamurai*. J Am Chem Soc 125: 15700-15701.

123. Wang J, Willumsen N, Zheng Q, Xue Y, Karsdal MA, et al. (2013) Bringing cancer serological diagnosis to a new level: focusing on HER2, protein ectodomain shedding and neoepitope technology. Future Oncol 9: 35-44.

124. Parks WC, Wilson CL, López-Boado YS (2004) Matrix metalloproteinases as modulators of inflammation and innate immunity. Nat Rev Immunol 4: 617-629.

125. Johnson PR (2001) Role of human airway smooth muscle in altered extracellular matrix production in asthma. Clin Exp Pharmacol Physiol 28: 233-236.

126. López-Boado YS, Wilson CL, Hooper LV, Gordon JI, Hultgren SJ, et al. (2000) Bacterial exposure induces and activates matrilysin in mucosal epithelial cells. J Cell Biol 148: 1305-1315.

127. Dahlen B, Shute J, Howarth P (1999) Immunohistochemical localisation of the matrix metalloproteinases MMP-3 and MMP-9 within the airways in asthma. Thorax 54: 590-596.

128. Daisuke Sakai (2008) Future perspectives of cell-based therapy for intervertebral disc disease. Eur Spine J 17: 452–458.

129. Christopher M Overall, Angela E King, Douglas K Sam, Aldrich D Ong, Tim TY Lau, et al. (1999) Identification of the tissue inhibitor of metalloproteinases-2 (TIMP-2) binding site on the hemopexin carboxyl domain of human gelatinase A by site directed mutagenesis. The hierarchical role in binding TIMP-2 of the unique cationic clusters of hemopexin modules III and IV. Journal of Biological Chemistry 274: 4421–4429.

130. Morgunova E, Tuuttila A, Bergmann U, Tryggvason K (2002) Structural insight into the complex formation of latent matrix metalloproteinase 2 with tissue inhibitor of metalloproteinase 2. Proc Natl Acad Sci U S A 99: 7414-7419.

Advances in Protein Chemistry

Chapter: Proteins and Peptides- Re-emergence in Prophylactics and Therapeutics

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published August, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Proteins and Peptides- Re-emergence in Prophylactics and Therapeutics

Heena Imtiaz Malik, Nazim Husain and Fahim H Khan*

Department of Biochemistry, Faculty of Life Sciences, Aligarh Muslim University, India

***Corresponding authors:** Fahim H Khan, Department of Biochemistry, Faculty of Life Sciences, Aligarh Muslim University, India, E-mail: fahimhalimkhan@hotmail.com

Abstract

Therapeutic peptides and proteins represent a rapidly growing part of marketed drugs and have an undisputed place alongside other established drugs .Peptide and protein drugs, classified as biopharmaceuticals or biodrugs, include monoclonal-antibody-based products for the treatment of different types of cancer and autoimmune diseases, therapeutic vaccines for immunization against HPV, cancer, hepatitis, insulin for diabetes treatment, human growth hormone for supplementation in hormone deficiency, and interferon a for treatment of various diseases and many more . The recent advances in large scale purification processes, recombinant DNA technology and analytical characterization have widened the horizons. Hopefully with benefits including favorable time to market and high rate of success in clinical trial compared with traditional pharmaceuticals, therapeutic proteins and peptides will play a pivotal role in the treatment of large number of diseases in the coming years.

Introduction

The field of proteomics and proteins is under extra-ordinary growth phase. This is largely due to the fact the human and other important genomes have been completely sequenced, which has opened the door for mining of wealth of information encoded in its sequence. This actively growing field has attracted scientists, biotechnologists and pharmaceutical companies that are taking keen interest in its development. Reason is proteins are the major functional molecules that are employed as catalysts, prophylactic and therapeutic agents apart from hundreds of other uses [1,2]. Recent research in the fields of molecular medicine has increased our understanding of diseases at the molecular level. This understanding allows scientists to deconstruct diseases as well as rationally design and develop its prophylactic and therapeutic approaches. Any breakthrough that occurs in any field of related to human health or has lasting impact on well being of human population is nothing short of miracle as it brings us closer to healthier, happier life

Protein/ peptide based therapeutics are largely constituted by antibodies, cytokines and small therapeutic peptides. Antibodies had always topped the list of therapeutic proteins in fact the half of the top 10 bestselling drugs in 2012 are monoclonal antibodies highlights the clinical importance of this novel class of therapeutics [2,3].This chapter is an attempt to make aware the students some of the applications in prophylaxis and therapeutics achieved by midas touch of proteins and peptides that are going to touch our lives more sooner than later.

HPV

Human Papillomaviruses (HPV) are common pathogens associated with a variety of benign and malignant epithelial lesions. HPVs can infect the squamous epithelia of genital tract can cause cancers of the cervix, vulva and vagina in women, cancers of the penis in men as well as anogenital warts in both sexes [4,5]. Infact, HPV infection is recognized as the necessary cause of cervical cancer. Cervical cancer is the second most common cancer in women and the fifth most common cancer overall. The high-risk types of human papillomavirus (HPV) are associated with over 80% of cervical cancers through epidemiological and experimental studies. Two high-risk types, HPV type 16 (HPV-16) and HPV type 18 (HPV-18), are responsible for up to 50% and 20% of all cervical cancers, respectively. There is no cure for HPV and successful preventive measures relies heavily upon its vaccination [6].

HPV is a non-enveloped, double-stranded, circular DNA virus with an icosahedral capsid [7]. Its genome is approximately 8 Kb long and is composed of early proteins (E1, E2, E4, E5, E6, E7) and late proteins (L1, L2). The functions of various genes and their products are highlighted in Figure 1. HPV enters the target cells by binding to its cellular receptor. This binding is dependent only on capsid protein L1 and does not require the other capsid protein L2. HPVs are generally internalized via a clathrin-dependent endocytic mechanism. After getting internalized and viral coat is disassembled which allow viral genome's access to the cellular transcription and replication machinery [8]. The virus infects the stem cells basal keratinocytes of the mucosal (genital) epithelium and delivers the genome to the nucleus. This episomal viral genome starts expressing the early proteins that maintain the keratinocytes in a rapidly dividing state [9,10]. Upon differentiation of infected cells, productive replication is established such that the viral genome is amplified to more than 1000 copies and expression of capsid proteins is induced, resulting in the synthesis of infectious virions that are assembled and released. The fact that the complete virus cycle takes place above the basal layer and without directly triggering cell lysis, limits the interaction of viral antigens(which do not spill out) with the immune cells of the host that keep an active surveillance for pathogens of the basal membrane [11]. Thus the immune system remains largely unaware of the infection that has taken place in the epithelium. Together these events help attain a sort of immune anergy that explains the general lack of local inflammation, the poor immune response to the virus, and the long persistence of HPVs even in healthy immunocompetent individuals [12].

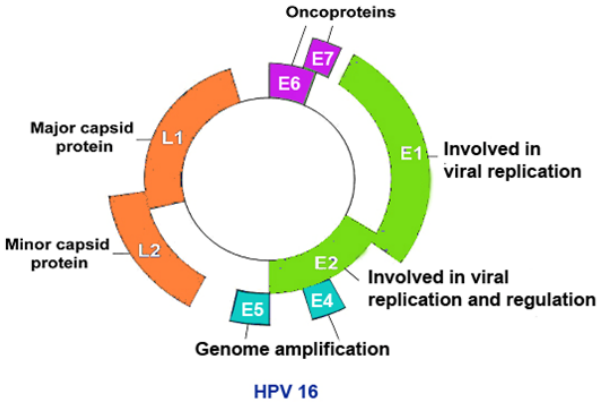


Figure 1: Schematic representation of the HPV-16 genome showing the location and functions of the early (E) and late genes (L1 and L2).

Treating HPV

To block the entry of HPV into target cell, prophylactic HPV vaccines aim to generate protective humoral immune responses in the host by stimulating the production of neutralizing antibodies against L1 and/or L2-HPV viral capsid proteins [13,14]. Antibodies formed against L1 and or L2 are neutralizing antibodies ie they prevent the entry of virion into the target cells, hence preventing the occurrence of infection. In manufacturing the vaccines, the viral capsid L1 gene is incorporated into a yeast genome or an insect virus genome using recombinant DNA technology (Figure 2). Grown in culture, the yeast or the insect cells produce the HPV L1 major capsid protein, which has the intrinsic capacity to self-assemble into virus-like particles (VLPs). These particles are subsequently purified for use in the vaccines. Currently, two HPV vaccines are available- First is a quadrivalent vaccine(Gardasil) effective against four HPV types 6, 11, 16, and 18 [15,16]. It comprises of VLPs of the four major mucosal HPVs: HPV type 16 (HPV- 16) and HPV-18 (the primary causes of cervical cancer) and HPV-6 and HPV-11 (the causes of genital warts).Second vaccine is (Cervarix), a bivalent vaccine against HPV types 16 and 18. The quadrivalent vaccine was approved by the US Food and Drug Administration (FDA) in 2006, and the bivalent vaccine was approved in 2009 [17]. Recombinant VLPs are morphologically indistinguishable from authentic HPV virions and contain the same type specific antigens present in authentic virions (Figure 3). Therefore, they are highly effective in inducing a host humoral immune response. And because they do not contain HPV DNA, the recombinant HPV vaccines(ie VLPs) are non-infectious and noncarcinogenic. Both vaccines have proven nearly 100% effective in protecting against the HPV types targeted by each vaccine in clinical trials that included hundreds of thousands of women and were fully licensed for use [17].

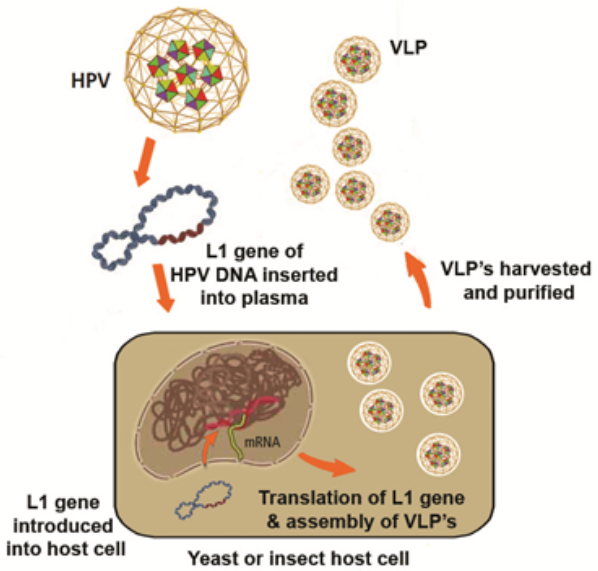


Figure 2: How HPV vaccines (VLP's) are produced.

Though its established that HPV vaccines are best way to prevent HPV infections and cervical cancer, the immunization program still faces many challenges, since HPV vaccination touches on issues related to adolescent sexuality, parental autonomy, and cost and because of that HPV immunization rates remain relatively low even in the United States. More recently, cytokines that act as immune-modulators with antiviral properties have been investigated for their role in HPV therapy [18]. An approved drug for HPV infections is interferon α (IFNα), a 19-kDa peptide/protein with antiviral properties against HPV. Interferon was the first FDA-approved immune-modulator for HPV. IFNs are cytokines with anti-viral, immune-regulatory and anti-proliferative properties. They have been used successfully in other viral infections and have been evaluated in the context of HPV-related disease. Three distinct types exist: IFN, IFN and IFN however for the treatment of HPV, IFN was found to be most effective. For HPV treatment, administration methods are restricted to intramuscular and intralesion injections, Recently, a topical cream was developed for the pain-free administration of IFNα that can be applied locally to infected areas and offers the advantages of providing sustained concentrations of the drug within affected skin while avoiding high systemic levels of the drug [16-18]. The topical cream is based on a delivery system called multi lamellar liposomes that encapsulate IFNα and transports the cytokine into skin and mucosa (Figure 4).

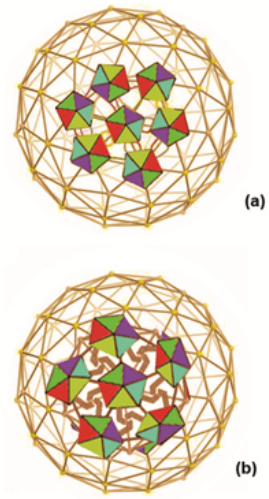


Figure 3: Schematic representation of human Papillomavirus – HPV (a) and virus like particle – VLP (b). Both HPV and VLP are of equal size and shape. VLP lack genome.

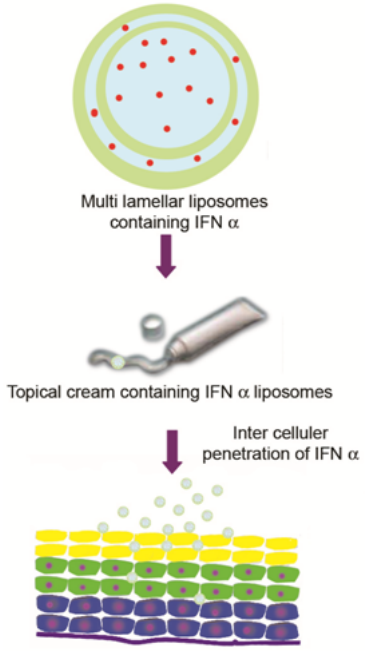


Figure 4: How topical cream containing IFN α acts.

H1N1 flu

The current outbreak of swine flu (in 2009) had been declared pandemic by world health organization .To effectively treat swine flu, antiviral medicines such as Oseltamavir (tamiflu) and zanamivir (relenza) needs to be administered within 48 hours of the onset of symptoms. Any delay in treating swine flu (or H1N1 flu) might risk patients life. So a quick detection test is the need of the hour. Pallane (Pallane medical Pvt. limited) has unveiled a virus detection test –RETCIF(Rapid Enhanced Tissue Culture Immunofluorescence) which is rapid and cost saving as compared to other viral detection tests that are currently used. This test, which was developed by Robert Alexander of Melbourne,Australia [19] ,uses five different cell lines, cell culture techniques, specific monoclonal antibodies and fluorescence microscope. Apart from detecting swine flu viruses , other respiratory viruses (such as respiratory syncyticial virus, adenovirus, influenza virus A,B, parainfluenza virus) cytomegalovirus ,herpes simplex virus can be detected in few hours .This test boasts of revealing the presence of virus in 3 hours and type of virus in 1-3 days.

RETCIF test involves taking patients sample (usually nasal swab in swine flu patients, broncoalveolar lavage, throat ,mouth ,skin swabs,cerebrospinal fluid, urine, faeces, nasopharygeal aspirates for other patients) and inoculating the sample on patented cell culture plates. These patented cell culture plates are 96 well microtiter plates that are inoculated with five different cell lines .Cell lines used are LLC-MK2,MDCK,Hep2,A549,MRC-5. Once viral samples are seeded onto cell culture plates, and supplemented with maintenance medium ,viruses are allowed to grow for limited time duration. Viruses are detected by adding specific monoclonal antibodies that detect specific viral proteins .This is followed by adding specific secondary antibody that binds this primary antibody .This secondary antibody is tagged with fluorophore that can be detected by fluorescence microscope. This viral diagnostic test RETCIF which is now commercially available, offers time saving procedure with greater accuracy than conventional shell vial method (which involves rapid cell culture).

Cancer

A normal cell undergoes regulated division, differentiation and cell death. When a cell has lost its control over their pathways of

normal division, differentiation and cell death (apoptosis) it becomes cancerous. So, cancer is the result of an abnormal proliferation of cells without differentiation and apoptosis. Cancer harms the body when cells divide uncontrollably to form lumps or masses of tissue called tumors. Tumors can be of two types benign and malignant. In benign tumor, cells remain clustered together in a single mass and do not spread to other sites [20]. More dangerous form of cancer is malignant tumor. In this form of tumour, cells don't remain localized at one site and becomes progressively invasive and spread to other tissues i.e become malignant. When a tumor successfully spreads to other parts of the body and grows, invading and destroying other healthy tissues, it is said to have metastasized. This process itself is called metastasis, and the result is a serious condition that is very difficult to treat. The term cancer refers specifically to malignant tumour.

Both benign and malignant tumors are classified according to the type of cell from which they arise. Most cancer fall into three major groups- carcinomas (tumors that arise from ectodermal or endodermal tissues), sarcomas (malignancies of mesodermal connective tissues) and leukemia or lymphomas (from blood forming tissues and cells of immune system).

Therapeutic approaches

Variety of therapeutic approaches are followed in the treatment of cancer including surgery, radiotherapy, chemotherapy, hormone therapy, anti angiogenesis therapy, and most common which is in scope of this chapter immune therapy (discussed below).Use of monoclonal antibodies, immune adjuvants, and vaccines against different form of cancers is referred to as immune therapy. Established therapies employ a variety of manipulations to either to kill tumour cells or to activate antitumor immunity. Collectively, these strategies attempt to augment protective antitumor immunity and to disrupt or kill the tumour cells or the immune-regulatory circuits that are critical for maintaining tumor. Since treatment of cancer by monoclonals is one of the most common and effective way of combating cancer, it is highlighted here.

Monoclonal antibodies: Antibodies are known to play an essential role in providing protective immunity to microorganisms, and the administration of tumor-targeting monoclonal antibodies has proven to be one of the most successful forms of immune therapy for cancer. The proper use of monoclonal antibodies is based on the fact that the tumour cell expresses antigen/antigenic determinants that are unique to that particular cell and not found on neighbouring healthy cells. This is due to the fact that transformation of normal cell to tumour cells is normally associated with expression of unique surface antigens on tumor cells that appear foreign to host immune system. Such antigens (called tumor antigens) are usually not expressed at all on normal cells, or expressed at very low concentration. Targeting of such antigens by monoclonal antibodies can have desired effect [20].

The infusion of preformed monoclonal antibodies can generate an immediate immune response while bypassing many of the limitations that impede endogenous immunity. The target specificity and optimal affinity of manufactured monoclonal antibodies can be carefully selected, and the quantity of antibody infused can be set to achieve biologically active antibody titers. Moreover monoclonal antibody therapies are typically not as toxic as conventional cytotoxic cancer chemotherapy. Many of the obstacles in initial use of murine monoclonal antibodies have been overcome by the generation of chimeric and humanised antibodies that contain less content of murine protein [21,22]. Murine proteins are generally considered foreign when administered to humans and hence immune response occurs that clears such proteins(including murine monoclonals) from the human system. Monoclonal antibodies consisting of mouse variable (Fab) regions and human constant (Fc) region are referred to as chimeric antibodies and hybrid antibodies that carry only murine CDRs in an otherwise human antibodies are referred to as humanized antibodies.

Several monoclonal antibodies, targeting tumor-associated proteins, are clinically approved for the treatment of cancer [23]. Five of these antibodies bind surface proteins that are highly expressed on tumors (Figure 5) namely CD52 in chronic lymphocytic leukemia (CLL) (alemtuzumab), CD33 in acute myelogenous leukemia (AML) (gentuzumab), and CD20 in non-Hodgkin's lymphoma (NHL) and CLL (rituximab, ibritumomab tiuxetan and tositumomab). Alemtuzumab (Campath) which is used to treat patients with CLL, cutaneous T-cell lymphoma (CTCL) and T-cell lymphoma. Alemtuzumab binds to CD52 antigen (present on the surface of mature B and T cells) and targets these cells for destruction. Bevacizumab is a recombinant humanized monoclonal antibody that binds to vascular endothelial growth factor (VEGF) and cetuximab, a chimeric antibody, targets epidermal growth factor receptor (EGFR). Intracellular signaling pathways related to VEGF and EGFR- both play a central role in tumorigenesis as well as tumor growth, and are therefore rational targets for anti-cancer drug development (Table 1). Targeting of such antigens by monoclonal antibodies can have desired effect i.e. they can be used to kill the target cell by antibody dependent cell mediated cytotoxicity (ADCC) (e.g in monoclonal antibodies targeted towards CD20 in follicular lymphoma in humans, induced ADCC) or complement mediated cell lysis (some antitumor effects of rituximab – a monoclonal antibody used in treatment of non-Hodgkin lymphoma are because of complement mediated cell lysis) (Figure 6). Complement dependent cytotoxicity constitutes a good tumor cell killing mechanism except solid tumor cells which have proved to be complement resistant.

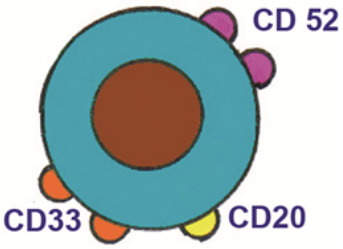


Figure 5: Some important cancer antigens that are targeted in therapy.

Monoclonal antibodies can also be used to deliver cytotoxic agents directly to tumors by conjugating them with radioactive isotopes or toxic chemicals, drugs or prodrugs [20,21,24]. Monoclonal antibodies attached to a radioactive substance, drug, or toxin, are referred to as conjugated monoclonal antibodies. The monoclonals are used as a targeting device to deliver these payloads directly to the cancer cells. Conjugated monoclonal antibodies circulate in the body until they can find and bind onto the target antigen on the tumour cell. It then delivers the toxic payload where it is needed most. This lessens the damage to normal bystander cells in other or neighbouring part of the body.

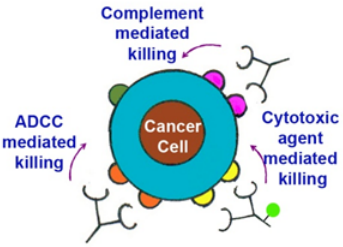


Figure 6: Different mechanisms of cancer cell killing by therapeutic antibodies ADCC- Antibody dependent cell mediated cytotoxicity.

Conjugated monoclonal antibodies can be broadly divided into two different types depending on what they are linked to. Monoclonals linked with chemotherapeutic drugs are referred to as chemolabeled.eg of one of the chemolabeled antibody approved by FDA is Transtuzumab emtansine (Monoclonal directed towards Her/neu receptor linked to Mertansine, a cytotoxic agent) for breast cancer [25]. Chemotherapeutic drugs such as doxorubicin and prodrug such as etoposide have been linked to monoclonals and selectively delivered to tumor site with some success [24,26].

Monoclonals linked with radioisotopes are referred to as radio labeled. Tositumomab (Bexxar) are examples of radiolabeled monoclonal antibodies. It is directed against the CD20 antigen and delivers its toxic payload directly to cancerous B cells. This antibody is used in treating some types of non-Hodgkin lymphoma [27]. Comprehensive list of monoclonal antibodies alongwith target antigens, indications and year of approval is given in Table 1.

Name	Target	Type	Approved for	Approval year
Muromonab-CD3	Anti-CD3	Murine	Prevention of kidney transplant rejection	1986
Abciximab	Anti-GPIIb/IIIa	Chimeric	Prevention of Platelet aggregation	1994
Rituximab	Anti-CD20	Chimeric	Non-Hodgkin's lymphoma	1997
Basiliximab	Anti-IL2R	Chimeric	Controlling kidney transplant rejection	1998
Daclizumab	Anti-IL2R	Humanized	Prevention of kidney transplant rejection	1997
Palivizumab	Anti-RSV	Humanized	Prevention of respiratory syncytial virus infection	1998
Infliximab	Anti-TNF	Chimeric	Crohn disease, rheumatoid arthritis	1998
Trastuzumab	Anti-HER2	Humanized	Breast cancer	1998
Gemtuzumab ozogamicin	Anti-CD33	Humanized	Acute myeloid leukemia	2000
Alemtuzumab	Anti-CD52	Humanized	Chronic myeloid leukemia	2001
Adalimumab	Anti-TNF	Human	Rheumatoid arthritis	2002
Tositumomab-I131	Anti-CD20	Murine	Non-Hodgkin lymphoma	2002
Efalizumab	Anti-CD11a	Humanized	Psoriasis	2003
Cetuximab	Anti-EGFR	Chimeric	Colorectal cancer, head and neck cancer, lung cancer	2004
Ibritumomab tiuxetan	Anti-CD20	Murine	Non-Hodgkin's lymphoma	2002
Omalizumab	Anti-IgE	Humanized	Allergic, Asthma	2003
Bevacizumab	Anti-VEGF	Humanized	Colorectal cancer, lung cancer	2004
Natalizumab	Anti-a4 integrin	Humanized	Multiple sclerosis, Crohn's diseases	2004
Ranibizumab	Anti-VEGF	Humanized	Macular degeneration	2006
Panitumumab	Anti-EGFR	Human	Colorectal cancer	2006
Eculizumab	Anti-C5	Humanized	Paroxysmal nocturnal hemoglobinuria	2007
Certolizumab pegol	Anti-TNF	Humanized	Crohn disease	2008
Golimumab	Anti-TNF	Human	Rheumatoid and psoriatic arthritis, ankylosing spondylitis	2009
Canakinumab	Anti-IL1b	Human	Muckle-Wells syndrome, rheumatoid arthritis	2009
Catumaxomab	Anti-EPCAM/CD3	Rat/mouse bispecific	Malignant ascites	2009
Ustekinumab	Anti-IL12/23	Human	Psoriasis multiple sclerosis	2009
Tocilizumab	Anti-IL6R	Humanized	Rheumatoid arthritis	2009
Ofatumumab	Anti-CD20	Human	Chronic lymphocytic leukemia	2009
Denosumab	Anti-RANK-L	Human	Osteoporosis	2010
Belimumab	Anti-BLyS	Human	Systemic lupus erythematosus	2011
Ipilimumab	Anti-CTLA-4	Human	Metastatic melanoma	2011
Brentuximab vedotin	Anti-CD30	Chimeric	Hodgkin lymphoma	2011
Pertuzumab	Anti-HER2	Humanized	Breast Cancer	2012
Raxibacumab	Anti-B. anthracis PA	Human	Anthrax infection, Prophylaxis and treatment	2012

BLyS: B lymphocyte stimulator; C5: complement 5; CD: cluster of differentiation; CTLA-4: cytotoxic T lymphocyte antigen 4; EGFR: epidermal growth factor receptor, EPCAM: epithelial cell adhesion molecule; GP: glycoprotein; IL: interleukin; NA: not approved; PA: protective antigen; RANK-L: receptor activator of NFkb ligand; RSV: respiratory syncytial virus; TNF: tumor necrosis factor; VEGF: vascular endothelial growth factor

Table 1: Brief overview of therapeutic monoclonal antibodies approved in the European Union, United States or Japan.

Cytokines: Cytokines, (secreted proteins with immune modulating properties), can be delivered systemically to activate antitumor immunity. Although response rates are low, both the cytokines IL-2 and IFN-α have been used to treat advanced melanoma and renal cell carcinoma, that are generally refractory to standard chemotherapy [26]. IFN-α is an important mediator in antiviral immunity, and IL-2 is a potent T cell growth factor, although how these functions contribute to the pharmacologic effect on tumour cells is presently unclear. Work in animal models suggests that IFN-α may play a role in antitumor immunity, linking the effectiveness of IFN-α to an induction of an immune response. The side effects of cytokine administration are severe and often dose limiting. Typically, cytokines induce symptoms that mirror those of systemic infection, including hypotension, vomiting, diarrhea, fever, and malaise. Another cytokine, IL-

21 increases the functional effects of NKT, NK and CD8+ T cells and may be effective in treating human cancers in combination with other first line therapies.

Autoimmune Diseases

Autoimmune disease is an immune malfunction in which immune response is directed against self molecules. They arise when the immune system responds aggressively to the individual's own tissues through recognition of self-antigens. Such an immune reaction which evokes pathological consequences could be due to involvement of humoral, cell mediated or complement mediated immunity. It is believed that autoimmune diseases affect 5-6 percent of the entire population. Among the most common autoimmune diseases are rheumatoid arthritis, multiple sclerosis, type-1 diabetes, pernicious anemia and Graves diseases which accounts for about 90 percent of all cases. In general, women are three times more likely than men to develop autoimmune diseases [27,28].

Lymphocytes specific for self-antigens are found in all individuals whether they are healthy or suffering from an autoimmune condition. The immune system maintains such broad responsive lymphocyte repertoire to provide protection against infections and cancers. However within this repertoire contains autoreactive cells but these cells are tightly suppressed and regulated by the immune system. Autoimmune diseases results from breakdown of such regulation or tolerance. Although there are varieties of mechanisms for induction of autoimmunity [20], the exact mechanism or causes remains to be worked out. The ideal treatment of autoimmune diseases would be induction of long lasting antigen specific tolerance. Achieving this mission is hindered by the fact that efforts to identify the specific provoking autoantigen in large number of auto-diseases are still underway. Moreover, induction of such a specific tolerance is difficult during ongoing immune response.

Therapeutic approaches

The current strategies for treating autoimmune diseases are aimed at providing management program that reduces pain and discomfort, prevents deformities and loss of bodily function so that patient leads a near normal life. The treatment strategies include treatment with therapeutic peptides (discussed below (Figure 7)), treatment with immunomodulatory and cytotoxic agents (azathioprine, cyclosporine, cyclophosphamide), non steroidal anti-inflammatoryagents (ibuprofen, aspirin, paproxen), corticosteroid (prednisone, methylprenisolone), disease modifying anti-rheumatic drugs (methotrexate, hydroxycholoroquine, sulfasalazine). These immunomodulatory and cytotoxic agents along with other drugs are either anti-inflammatory or immunosuppressive and are already widely used for autoimmune as well as other diseases though with some (sometimes life threatening) side effects.

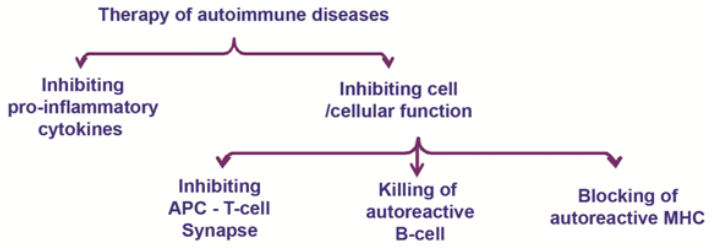


Figure 7: Protein/Peptide based therapy of autoimmune diseases.

There are number of therapeutic approaches that have been used to treat or manage autoimmunity in experimental animal models and human clinical studies [20]. These therapies usually involve proteins such as antibodies, peptides such as cytokines that target either the culprit causative agents (eg cells) or eliminated cytokines that induce pro inflammatory reactions. Soluble peptides such as glatiramer [29] serve as tolerogen when delivered inside the host body. Such peptides tend to induce "tolerance" in the host which results in depression in auto immune reactions. Use of some important proteins and peptides in managing autoimmune diseases are discussed below.

Targeting cytokines: New and novel non-antigen specific therapeutic approaches target cytokines. As mentioned previously, cytokines are protein mediators of inter-cellular communication and are involved in a wide variety of biological activities including immune responses. Role of different cytokines in autoimmune diseases have been well studied and it has become clear that excessive production of pro-inflammatory cytokines as well as lack of anti-inflammatory ones constitutes to pathogenesis of several autoimmune diseases. Fine tuning the cytokine balance may have therapeutic value and can be achieved by neutralization of pro-inflammatory cytokines or administering or inducing anti-inflammatory ones.

There are three major pro-inflammatory cytokines – Tumor necrosis factor α (TNF-α), interleukin-1 (IL-1) and IL-6. These cytokines can be *neutralized* in a variety of different ways such as production of soluble cytokine receptor, monoclonal antibodies against cytokines or production of non-activating cytokine receptor antagonists [30]. Clinical trial with chimerized monoclonal antibodies against TNF (infliximab) in human have given encouraging results in rheumatoid arthritis (RA), Crohn disease (CD),psoriasis and psoriatic arthritis [31]. Infliximab (trade name -remicade) is a chimeric monoclonal antibody against TNF that has antigen binding site of mouse origin and remaining antibody derived from human IgG1 antibody. After infliximab administration, clinical results were encouraging even when RA patients were resistant to multiple DMARDs (disease modifying anti-rheumatic drugs). Patient having steroid-resistant active CD also showed clinical and endoscopic improvement by single infusion of infliximab.

Similar results were reported for fully human anti-TNF antibody –Adalimuab (humira) that was found to be effective in RA, CD, and psoriatic arthritis.

IL-1, another proinflammatory cytokine have been implicated in the pathogenesis of autoimmune diseases such as RA [31]. Neutralizing of IL-1 have been achieved by IL-1 receptor antagonist which blocks the binding of both IL-1α and IL-1β to IL-1 receptor. Anakira, a human recombinant IL-1 receptor antagonist is approved for the treatment of RA. Blocking of other inflammatory cytokines such as IL-6, IL-15 have also proven successful strategy in ameliorating RA. Neutralizating IL-6, a cytokine that upregulates secretion of acute phase proteins by humanized monoclonal, anti IL-6 receptor antibody in rheumatoid arthritis are currently in progress.

Systemic administration of recombinant cytokines to patients of autoimmune diseases have also been explored with some success

[30,33]. Clinical trials with autoimmune skin disease –psoriasis showed marked improvement with the administarion of IL-4. Similar administration of cytokine IL-11, regarded as anti-inflammatory and associated with immuno-modulating functions is currently being explored in patients with CD.

IFN-β has showed potential in treatment of multiple sclerosis and management of this disorder. Administering IFN-β reduced relapse rate of multiple sclerosis (MS) in patients with relapsing/remitting MS by 30 per cent. IFN-β also delayed progression for about a year in patients with secondary progressive multiple sclerosis [33].

IFN-α expression is elevated in SLE and type-I diabetes and biological effects of IFN-α can largely, if not completely account for its pathology. Due to lack of curative therapies for these disease, decreasing IFN-α concentration for a specific period of time is one of the possible strategy. Humanized monoclonal antibody directed against IFN-α is currently being explored as one agents that could decrease IFN-α in SLE patients [33].

Targeting cells: T cells-New enthusiasm is being generated for agents that are known as T cell co-stimulatory blockers. These blockers inhibit constructive synapse between T cell and antigen presenting cells. This immunologic synapse is formed between CD80/CD86 found on antigen presenting cells and CD28 found on T cells. This synapse between antigen presenting cell and T cell is blocked by fusion protein *cytotoxic T lymphocyte associated antigen 4* IgG1(CTLA-4Ig) which binds CD80/CD86 with high affinity and prevents the formation of constructive synapse between these two cells. This T cell costimulatory blocker –Abtacept (tradename-Orencia) has shown to improve symptoms of RA [34]. It should be remembered that it is only after the formation of constructive synapse (which means ligation of T cell receptor with antigen MHC complex and binding of CD80/CD86 with CD28) that T cell becomes active and start proliferating, and secreting proinflammatory cytokines (such as TNF) which harms the host. When Abtacept binds CD28 present on T cell surface,it blocks co stimulatory signal from being delivered from antigen presenting cell to T cell. This turns down T cell response which results in decreased production of TNF.

B cells-Lot of interest is being generated these days from clinical effectiveness of depleting B cells with antibodies that bind CD20 present on B cell surface. This leads to the removal of B cells from the circulation. B cell serves as both antigen presenting cell as well as antibody forming cell. Removal of B cells has been shown to be effective in reducing autoantibody levels in RA and clinical improvement have been observed in majority of patients in RA. One B cell depleting agent –Rituximab (tradename-Rituxan), a chimeric monoclonal antibody against CD20 is currently available for treatment of RA [35].

Targeting MHC: Glatiramer acetate (also known as Cop-1, or Copaxone) is one the most widely prescribed drug that is currently used to treat multiple sclerosis [29]. It is a random polymer of four amino acids found in myelin basic protein (MBP) namely lysine, alanine, glutamic acid, tyrosine and has a average molecular mass 6.4 kDa. It is suggested that when this drug is administered, it competes with MBP epitope binding to disease specific MHC that is present in diseased individual. In other words it blocks the binding of "rogue". MBP peptides to its corresponding MHC which results in decreased activation of MBP specific clones of T cell. In mouse models, Glatiramer proved very effective as it polarised cytokine profile towards Th2 cells that suppresses the inflammatory response.

Food Allergy

Food allergy is an adverse health effect that results from a specific immune response that occurs repeatedly on exposure to a given food. A food is defined as any substance intended for human consumption, including drinks, food additives, and dietary supplements. All adverse food reactions be catogorized as either immune-mediated (food allergy and celiac disease) or nonimmune mediated (formerly known as food intolerances). The immune-mediated adverse food reactions can be immunoglobulin E (IgE)-mediated or non–IgE-mediated [20,36].

The mechanisms of immune-mediated adverse food reactions involve recognition of harmless antigens of food by the cells of adaptive immunity, the T and B lymphocytes. IgE mediated reactions involve differentiation of naive T cells into Th2 cells [36,37]. The Th2 cells stimulate the B lymphocytes to produce allergen-specific IgE antibodies; these attach themselves to the surfaces of mast cells and blood basophils by way of a receptor, a process known as sensitization. When recognition of an allergen by the specific IgE antibodies occurs, the mast cell de-granulates, releasing histamine, prostaglandins, and leukotrienes that provoke the characteristic symptoms of IgE-mediated food allergy.

Non–IgE-mediated food allergic responses are not our primary concern here though they usually involve differentiation of naïve T cells into Th1 cells, or induction of activation of other cells of the immune system such as eosionophils. The pathogenic mechanisms underlying the reactions of nonimmune mediated adverse food reactions are diverse, and some remain unknown.

Therapuetic approaches

Milk, egg, and peanut account for the vast majority of food-induced allergic reactions in American children, whereas peanut, tree nuts, fish, and shellfish account for most of the food-induced allergic reactions in American society [38,39]. Food allergy affects up to 6% of children and 3–4% of adults in Westernized countries. The best method of treatment of food allergy is strict avoidance of the offending allergens and awareness regarding the use of emergency medication in cases of accidental ingestions or exposures. While these approaches are generally effective, there are no definitive treatments that provide for long-term remission from food allergy. However, with recent advances in characterizing food allergens and understanding humoral and cellular immune responses in food allergy, several therapeutic strategies are being investigated. Potential treatments include allergen-specific immunotherapy as well as allergen-nonspecific approaches that diminish the overall allergic response in food-allergic individuals.

Immunotherapy: A major component of immunotherapy appears to be the alteration of the T-cell responses to the allergen via skewing the proallergic Th2 response towards a Th1 response that does not cause allergic reactions. Below we briefly mention some of the approaches that are used in immune therapy of food allergies and involve use of proteins and peptides (Figure 8).

The generation of 'protective' antibody responses: Desensitization is one the method through which protective antibody can be generated. It is a method to reduce or eliminate the hosts adverse reaction to a substance. Person is desensitized by administering miniscule doses of allergen so that allergic individuals develop protective IgG and not offending IgE. For example, if a person with allergy to food product (eg: peanut) has an adverse allergic reaction when he consumes it, the doctor may give the person a very small amount of the food product at first. Over a period, larger doses are given until the person is taking the full dose. This is one way to help the body get used to the full dose and to avoid having the allergic reaction to food product. At the cellular level, administration of small doses of food

allergen induces IgG response which eventually overrides the allergic IgE response [39,40]. It is an interesting approach although it is not effective all the time. Peanut allergy has attracted a lot of attention, since very severe reactions can occur with it, which makes accidental ingestions more worrisome. One study reports a 38-year-old woman with peanut allergy who underwent oral desensitization. At the end of desensitization and maintenance phase of 6 months she can eat 40 g of peanuts three-times a week daily for an indefinite period of time; moreover, she was able to tolerate peanut-containing products without issues.

Figure 8: Immuno therapy for food allergy.

Administration of heat-denatured proteins: This approach of immunotherapy revolves around the fact the allergic antibody binds to both sequential and conformational epitopes on allergic proteins. Heat denaturation destroys conformational epitopes and makes allergic protein tolerable [41]. Studies on cow's milk have shown that, children that were allergic to milk, 73% tolerated extensively heated milk products. Similarly another group saw similar results in children with hen's egg allergy being challenged with baked egg products. This may represent another avenue to explore with regard to therapy of food allergy.

Peptide immunotherapy: One of the major drawbacks of protein based immunotherapy is the potential for severe systemic reactions, since whole allergens(which are usually large molecules) are associated with the risk of inducing allergic reaction owing to the presence of B-cell epitopes on the allergen that can crosslink IgE [36,37]. Peptide-based immunotherapy uses short peptides instead of large proteins. It can prevent the crosslinking of IgE and hence triggering mast cells and basophils which mediate allergic reaction. This was studied regarding cat hyper-sensitivity, with significant clinical improvement in symptoms. Studies have yet to reach the clinical trial stage in peptide immunotherapy, but this may be a future direction.

Use of recombinant allergens: This concept involves engineering antigenic epitope-free proteins through PCR mutagenesis and as in peptide immunotherapy, allows an antigen to influences specific T cells without crosslinking IgE. This has been investigated in the case of peanut allergy, where mutated Ara h 1, 2 and 3 (HKE-MP123) have been engineered to be expressed by Escherichia coli and then administered in murine model [42]. The outcome demonstrated a decrease in food challenge reactivity compared with the native allergen.

Anti-IgE therapy: Anti-IgE is a humanized mouse monoclonal IgG antibody directed against the human IgE molecule which binds to freely circulating IgE, thereby rendering it unable to bind to its specific high-affinity receptor (FcεRI) on mast cells and basophils. This works through an 'allergen nonspecific' approach, whereby it can be effective against multiple different allergens. AntiIgE has already been studied and administered successfully in asthma and allergic rhinitis. Given the success of these studies, humanized monoclonal anti-IgE [43] has recently been used in animal studies and preliminary clinical studies in human with considerable success.

Treatment type	Key features	Clinical trials on Human
Desensitization	Incrementally increased swallowed allergen Dose resulting in desensitization, Tolerance	Yes
Heat-denatured protein	Extensively heated foods may be less Allergenic	Yes
Peptide immunotherapy	Prevents IgE crosslinking, mast cells not activated	No
Anti-IgE immunotherapy	Decreases IgE available for allergen binding	Yes

Table 2: Immune therapies for food allergy.

In the past decade, new expression systems have emerged as a serious competitive force in the large-scale production of recombinant proteins and peptides.Some products of these expression systems have reached market and is likely to become Gen-next in production in the proteins and peptides. Very briefly, we discuss two important ones.

Molecular Pharming

Pharming is derived from two words- farming and pharmaceuticals. It refers to large scale production of recombinant pharmaceutical proteins in which living cells or organisms (such as plants) are used as expression hosts. In simple words, pharming or molecular farming as it is often called, involves the use of genetically enhanced plants to produce varied array of pharmaceutical proteins (Figure 9). Growing recombinant proteins in plants is relatively new concept and it has evolved as a method that is simple, inexpensive and allow large scale production of safe recombinant proteins [44,45]. Until recently, most plant derived proteins have been produced in transgenic tobacco plants and extracted directly from leaves. These proteins were produced in low levels typically less than 0.1 % of the total soluble proteins. As an alternative to transgenic plants, transplastomic plants are now being used to achieve higher yields. Transplastomic plant is a transgenic plant in which the desired transgene is introduced into the plastid genome rather than nuclear genome by the process of particle bombardment. Transplastomic plants provide higher yield of expressed transgene which in some cases can be as high as 25% of the total soluble protein [45]. Other plants that nowadays used in molecular farming include maize (for producing antibodies, enzymes such as trypsin, laccase, proteins such as avidin),vegetables such as potatoes (for antibodies,human milk protein), tomatoes (for producing rabies vaccine) as well as other fruit plants such as banana which can be consumed raw by adult and children. As can be seen a large number of proteins ranging from enzymes to complex proteins such as antibodies have been expressed in plants but we shall focus on only those proteins that come under our purview and have diagnostic, therapeutic and prophylactic applications.

Proteins currently being produced in plants for molecular farming purposes can be categorized into two main categories areas: (1) Therapeutic proteins (antibodies,vaccine peptides etc) (2) industrial proteins (e.g., enzymes). Industrial proteins are out of the scope of this chapter and will not be discussed.

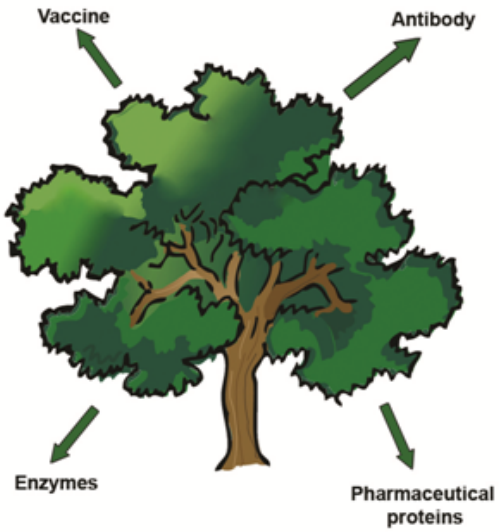


Figure 9: Molecular Pharming – Some important products of molecular pharming are depicted here.

Therapeutic Proteins

This group includes all proteins used directly as pharmaceuticals along with those proteins used in the making of pharmaceuticals. The list of such proteins is long, growing, and includes such products as thrombin and collagen, trypsin and aprotinin, antibodies (IgA, IgG, IgM, secretory IgA, etc.) and antibody fragments [46]. These can be produced in plants in both glycosylated and nonglycosylated forms. The plant-derived monoclonals (plantibodies) have the potential of alleviating the serious production bottleneck that currently exists as dozens of new monoclonals products attempt to reach the marketplace.

Antibodies were initially produced in tobacco plants but now cereal seeds are preferred as protein (antibody) accumulation in dry seeds allow long term storage at ambient temperature without any significant loss of activity. At least six types of plant derived recombinant antibody have entered preclinical testing stage [44-46]. The antibodies are (a) *Avicidin*, IgG molecule that recognizes adhesion molecule –EpCAM which is marker of colorectal cancer. This plant derived antibody was first to be administered on humans. It was produced in corn.(b) *CaroRx* is a chimeric IgG/IgA antibody that recognizes main adhesion protein of *Streptococcus mutans*, a pathogen for tooth decay in humans, is presently in phase II clinical trials. (c) *T*84.66 is monoclocal antibody that is expressed in plants that recognizes carcinoembryonic antigen, a well know marker of epithelial cancers. Other plant derived pharmaceutical antibodies that have achieved commercial status include antiherpes simplex virus antibody (that recognizes herpes simplex virus -2 glycoprotein, has been expressed in soybean plant). Anti respiratory syncytial virus antibody (this IgG type antibody is being developed by a United States company Epicyte Pharmaceuticals, recognizes R9 protein of respiratory syncytial virus), 38C13 antibody (it is unique antibody that directed towards mouse lymphoma cells.Efforts are to develop this system as an effective therapy for human diseases such as non Hodgkin lymphoma), PIPP (a monoclonal antibody that recognizes human chorionic gonadotropin (hCG), that could be used in diagnosis and/ or therapy of tumors that produce hCG, as well as in pregnancy detection) are undergoing clinical trials [47].

The concept of oral vaccine with vaccine antigens expressed in raw fruits vegetables, seeds has risen to popularity. Two other pharmacological important antigens have been expressed in potato and they are entering clinical trial phase [46,47]. These vaccine antigens are heat labile toxin B (LT-B) subunit of E.coli and capsid protein of Norwalk virus. These two antigens from two very important enteric pathogens,and hence constitute two ideal oral vaccine candidates. In fact clinical trials of LT-B showed that consumption of raw potato that contained about 0.5 -10 mg of LT-B produced high titer of mucosal and systemic antibodies. Other proteins of medical importance that have been produced in plants include collagen(tobacco), proteinase inhibitor (rice), milk protein protein-caesin(potato), erythropoietin(tobacco), human serum albumin(potato), Rabies virus glycoprotein (tomato), diabetes autoantigen (tobacco, potato), porcine glycoprotein gastroenteritis virus glycoprotein S (tobacco, maize), lysozyme and protein polymers that can be used in surgery and tissue replacement. The fact that plant derived avidin and β-glucuronidase are now being marketed by Sigma Inc at lower price shows the potential of molecular farming in plants to compete with established market and technology [47].

Therapeutics Involving Peptides

Over the years the peptides are not generally considered to be a good drug candidate because of their low bioavailability and propensity to be rapidly metabolized [48]. New strategies to improve productivity and reduce metabolism of peptides have been developed in recent years, and a large number of peptide-based drugs are now being marketed.

Therapeutic peptides traditionally have been derived from four sources- natural peptides produced by plants, animal or human origin (eg. derived from naturally occurring peptide hormones), peptides isolated from genetic or recombinant libraries and peptides discovered from chemical libraries. Bioavailability of peptides drug are determined by number of factors such as absorption, transport, and passage across biological membranes [49]. When compared with therapeutic proteins and antibodies, peptide drug candidates do have notable drawbacks: they generally have low stability inplasma, are sensitive to proteases and can be cleared from the circulation in a few minutes. However the main limitations generally attributed to therapeutic peptides are: low oral bioavailability, a short half-life; and poor ability to cross physiological barriers [48,49]. The greatest threat to therapeutic peptides comes from enzymes present in various parts of the body for example the lumen of the small intestine, which contains gram quantities of peptidases secreted from the pancreas (e.g. achymotrypsin, trypsin, pancreatic elastase, carboxypeptidases A, as well as cellular peptidases from mucosal cells [49,50]. Another enzymatic barrier is the brush border membrane of the epithelial cells, which contains more than dozen peptidases e.g. dipeptidyl-peptidase IV, prolyl tripeptidylpeptidase, angiotensin-converting enzyme, leucyl-aminopeptidase, aminopeptidase M, aminopeptidase A, neprilysin etc), that together have a broad specificity and can degrade both peptides and proteins [50]. Proteolytic peptide degradation, results in short half-life (generally, a few minutes or, at best, a few hours). These shortcomings are now overcome by using new synthetic

strategies that introduces chemical modifications into peptides containing natural amino acids that limits proteolytic degradation, and using alternative routes of administration(injection) [49,51].

Keeping aside these shortcomings which we have almost overcome it should be noted peptides have the potential to penetrate further into tissues owing to their smaller size. Moreover, therapeutic peptides, even synthetic ones, are generally less immunogenic than recombinant proteins and antibodies [51,52]. Peptides have other advantages over proteins and antibodies asdrug candidates, including lower manufacturing costs, 15-50 higher activity per unit mass, greater stability (long storage at room temperature a), and better organ or tumour penetration. Therapeutic peptides also offer several advantages over small organic molecules that make up traditional medicines. The first advantage is that often representing the smallest functional part of a protein, they offer greater efficacy, selectivity and specificity than small organic molecules. A second advantage is that the degradation products of peptides are amino acids, which minimizing the risks of systemic toxicity. Third, because of their short half-life, few peptides accumulate in tissues (reduction of risks of complications caused by their metabolites). As result of these advantages, large number of therapeutic peptides and proteins have mushroomed over years [51] . Table 3 provides brief overview of some important proteins and peptides that currently used as drugs and table 4 summarizes some of important peptides and proteins that are currently in clinical trials.

Protein/peptide	Indications
Peptides	
• Corticorelin	Diagnosis of cushings syndrome
• Lisinopril	Hypertension
• Exenatide	Controls type 2 diabetes mellitus
• Human calcitonin	Pagets disease, postmenopausal osteoporosis
• Epitifibatide	Acute coronary syndrome
• Icatibant acetate	Hereditary angioedema
• Pramlintide	Both type 1 and type 2 diabetes
• Buserelin	Prostrate cancer
• Goserelin	Breast cancer
• Octreotide	Acromegaly
Proteins	
Growth hormone	Growth hormone deficiency
Human serum albumin	Livea cirrhosis, burns
Lactoferrin	Antimicrobial activity
Hirudin	Thrombin inhibitor
IgG/IgM/IgA	Therapeutic
Hepatitis B virus envelope protein	Vaccine
Rabies virus antigen	Vaccine (edible)
Norwalk virus antigen	Vaccine
Protein C	Anti coagulant
Collagen	Therapeutic/cosmetic
Erythropoietin	Anemia

Table 3: Important protein and peptide that have reached American European union or Japanese market.

Product	Type	Tested for
Anti IL-13 antibody	Antibody	Therapy, also used for assays, flowcytometry
Adnectin	High affinity Ig-like protein	Enginnered domain of fibronectin, used as bispecific antibody
Pegdinetanib	Antagonist of growth factor receptor	Anti cancer drug
gp 100	Protein	Vaccine for melanoma
HPV – 16/E7	Protein	Vaccine for cervical cancer
Vedolizumab	Antibody	Crohn disease
Trastuzumab	Antibody	Brest cancer
Corticorelin acetate	Peptide, corticotrophin releasing factor	Anti tumor activity

Table 4: List of some important proteins and peptides that are currently in clinical and preclinical trials.

Bicyclic Peptides

The poor tissue penetration, difficult and expensive procedure for chemical synthesis, and the need for application by injection led scientists to wonder about other alternatives such as smaller protein scaffolds as alternatives to antibodies. Polycyclic peptides could deliver better tissue penetration, higher activity per mass and a wider choice of administration routes than antibodies. Christian Heinis who was studying, protein-based scaffolds discovered that the bicyclic peptide structure could interact with the protein much like antibodies i.e, they act as small, adaptable antibodies that can bind to antigens via their surface loops. Such peptides could potentially be chemically synthesised, and at least have equivalent binding qualities to antibodies [53].

In 2012, Winter and Heinis formed a company Bicycle Therapeutics, which is based in Cambridge, UK. The prime focus of this company is on medical conditions for which the use of bicyclic peptides promises advantages over both small molecules and monoclonal antibodies. They have recently engineered potent bicyclic peptide (Figure 10) inhibitors that specifically block human kallikrein and murine proteases, but not other proteases. The high selectivity ensures that drug (peptide) binds to target only and to multitude of other binding sites as the nonspecific binding can often lead to toxicity [53].

Our increasing understanding of biological sciences will continue to define new targets for immune therapies and advancements of protein sciences will provide us with vision and tools for a safer and effective therapy. Hopefully, some of the recent breakthroughs in protein biology will advance science, accelerate technology and discoveries, affecting varied range of diseases. We all hope that these discoveries will flourish and provide hope, enrich our lives, create dreams and help us rejoice in this exciting period of scientific discovery.

Coming years will witness increasing number of exciting therapeutic trials. We will also encounter undesirable surprises along the way but with the zeal and success seen in past few years we can be sure that era of pathogenesis based therapy is there in not so distant future.

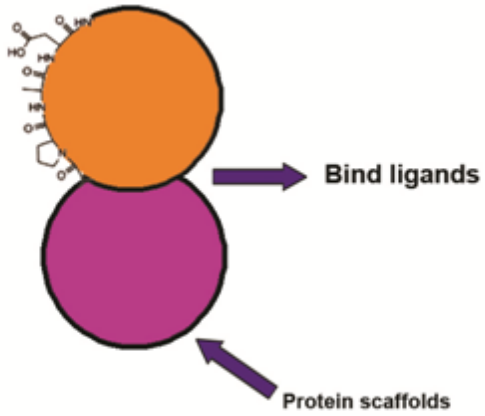


Figure 10: Schematic representation of bicyclic protein.

Acknowledgement

The authors wishes to thank Dr. Haseeb Ahsan , for critically reading the manuscript and giving helpful suggestions.

References

1. Rawlins MD (2004) Cutting the cost of drug development? Nat Rev Drug Discov 3: 360-364.
2. Di Masi J A and Grabowski H G (2007) The cost of biopharmaceutical R&D is biotech different Manag. Decis. Econ 28: 469–479.
3. http://www.prweb.com/releases/2013-monoclonal-antibody/industry analysis-report/
4. zur Hausen H (1987) Papillomaviruses in human cancer. Appl Pathol 5: 19-24.
5. Dunne EF, Unger ER, Sternberg M, McQuillan G, Swan DC, et al. (2007) Prevalence of HPV infection among females in the United States. JAMA 297: 813-819.
6. Ramet J, van Esso D, Meszner Z; European Academy of Paediatrics Scientific Working Group on Vaccination (2011) Position paper--HPV and the primary prevention of cancer; improving vaccine uptake by paediatricians. Eur J Pediatr 170: 309-321.
7. Mallik H I, Khan F H, Ahsan H (2013) Human Papilloma virus- Current status and issues of vaccination. Archives of Virology. In press
8. Garcea RL, Chen X (2007) Papilloma virus structure and assembly. In: R L Garcea, D. DiMaio (eds.). The papilloma viruses. Springer, New York, USA.
9. Danos O, Engel LW, Chen EY, Yaniv M, Howley PM (1983) Comparative analysis of the human type 1a and bovine type 1 papillomavirus genomes. J Virol 46: 557-566.
10. Moody CA, Laimins LA (2010) Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer 10: 550-560.
11. Grabowska AK, Riemer AB (2012) The invisible enemy - how human papillomaviruses avoid recognition and clearance by the host immune system. Open Virol J 6: 249-256.
12. Ashrafi GH, Brown DR, Fife KH, Campo MS (2006) Down-regulation of MHC class I is a property common to papillomavirus E5 proteins. Virus Res 120: 208-211.
13. Reiter PL, McRee AL, Kadis JA, Brewer NT (2011) HPV vaccine and adolescent males. Vaccine 29: 5595-5602.
14. Fraser C, Tomassini JE, Xi L, Golm G, Watson M, et al. (2007) Modeling the long-term antibody response of a human papillomavirus (HPV) virus-like particle (VLP) type 16 prophylactic vaccine. Vaccine 25: 4324-4333.
15. Joura EA, Leodolter S, Hernandez-Avila M, Wheeler CM, Perez G, et al. (2007) Efficacy of a quadrivalent prophylactic human papillomavirus (types 6, 11, 16, and 18) L1 virus-like-particle vaccine against high-grade vulval and vaginal lesions: a combined analysis of three randomised clinical trials. Lancet. 369: 1693-1702.
16. Paavonen J, Jenkins D, Bosch FX, Naud P, Salmerón J, et al. (2007) Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with human papilloma virus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. Lancet. 369: 2161-2170.
17. Mello MM, Abiola S, Colgrove J (2012) Pharmaceutical companies' role in state vaccination policymaking: the case of human papillomavirus vaccination. Am J Public Health 102: 893-898.
18. Beutner KR, Ferenczy A (1997) Therapeutic approaches to genital warts. Am J Med 102: 28-37.
19. Alexander R, Hurt AC, Lamb D, Wong FY, Hampson AW, et al. (2005) A comparison of a rapid test for influenza with laboratory-based diagnosis in a paediatric population. Commun Dis Intell Q Rep 29: 272-276.
20. Khan FH (2009) The Elements of immunology. Pearson Education India, India.
21. Morrison SL, Johnson MJ, Herzenberg LA, Oi VT (1984) Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. Proc Natl Acad Sci U S A 81: 6851-6855.
22. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G (1986) Replacing the complementarity-determining regions in a human antibody with those from a mouse. Nature 321: 522-525.
23. Cheung NV (2008) Therapeutic antibodies and immunologic conjugates. In: Abeloff, Doroshow, Tepper, Armitage, Niederhuber, et al. (eds.). Abeloff's Clinical Oncology. (4thedn), Churchill Livingstone, Philadelphia, USA.
24. Hudson PJ, Souriau C (2003) Engineered antibodies. Nat Med 9: 129-134.
25. Hudis CA (2007) Trastuzumab--mechanism of action and use in clinical practice. N Engl J Med 357: 39-51.

26. Walczak RJ, Dickens ML, Priestley ND, Strohl WR (1999) Purification, properties, and characterization of recombinant Streptomyces sp. strain C5 DoxA, a cytochrome P-450 catalyzing multiple steps in doxorubicin biosynthesis. J Bacteriol 181: 298-304.

27. Encyclopedia of immunology (1998) P J Delves (eds.). (2ndedn), Science direct, Amsterdam, Netherlands.

28. Davidson A, Diamond B (2001) Autoimmune diseases. N Engl J Med 345: 340-350.

29. Stüve O, Youssef S, Weber MS, Nessler S, von Büdingen HC, et al. (2006) Immunomodulatory synergy by combination of atorvastatin and glatiramer acetate in treatment of CNS autoimmunity. J Clin Invest 116: 1037-1044.

30. Hata H, Sakaguchi N, Yoshitomi H, Iwakura Y, Sekikawa K, et al. (2004) Distinct contribution of IL-6, TNF-alpha, IL-1, and IL-10 to T cell-mediated spontaneous autoimmune arthritis in mice. J Clin Invest 114: 582-588.

31. Maini R, St Clair EW, Breedveld F, Furst D, Kalden J, et al. (1999) Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet 354: 1932–1939.

32. Feldmann M, Brennan FM, Maini RN (1996) Role of cytokines in rheumatoid arthritis. Annu Rev Immunol 14: 397-440.

33. O'Shea JJ, Ma A, Lipsky P (2002) Cytokines and autoimmunity. Nat Rev Immunol 2: 37-45.

34. Sharpe AH, Abbas AK (2006) T-cell costimulation--biology, therapeutic potential, and challenges. N Engl J Med 355: 973-975.

35. Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, et al. (2006) Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. Arthritis Rheum., 54: 2793-2806.

36. Chaplin DD (2006) 1. Overview of the human immune response. J Allergy Clin Immunol 117: 430-435.

37. Lin J, Sampson HA (2009) The role of immunoglobulin E-binding epitopes in the characterization of food allergy. Curr Opin Allergy Clin Immunol 9: 357-363.

38. Savage JH, Matsui EC, Skripak JM, Wood RA (2007) The natural history of egg allergy. J Allergy Clin Immunol 120: 1413-1417.

39. Sicherer SH, Muñoz-Furlong A, Godbold JH, Sampson HA (2010) US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. J Allergy Clin Immunol 125: 1322-1326.

40. Patriarca G, Nucera E, Roncallo C, Pollastrini E, Bartolozzi F, et al. (2003) Oral desensitizing treatment in food allergy: clinical and immunological results. Aliment Pharmacol Ther 17: 459-465.

41. Husby S (2000) Normal immune responses to ingested foods. J Pediatr Gastroenterol Nutr 30: 13-19.

42. Li XM, Srivastava K, Grishin A, Huang CK, Schofield B, et al. (2003) Persistent protective effect of heat-killed Escherichia coli producing "engineered," recombinant peanut proteins in a murine model of peanut allergy. J Allergy Clin Immunol 112: 159-167.

43. Milgrom H, Fick RB Jr, Su JQ, Reimann JD, Bush RK, et al. (1999) Treatment of allergic asthma with monoclonal anti-IgE antibody. rhuMAb-E25 Study Group. N Engl J Med 341: 1966-1973.

44. Stöger E, Vaquero C, Torres E, Sack M, Nicholson L, et al. (2000) Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies. Plant Mol Biol 42: 583-590.

45. Fischer R, Emans N (2000) Molecular farming of pharmaceutical proteins. Transgenic Res 9: 279-299.

46. Giddings G (2001) Transgenic plants as protein factories. Curr Opin Biotechnol 12: 450-454.

47. Andersen DC, Krummen L (2002) Recombinant protein expression for therapeutic applications. Curr Opin Biotechnol 13: 117-123.

48. Marx V (2005) Watching peptide drugs grow up. Chem. Eng. News 83: 17–24.

49. Sato AK, Viswanathan M, Kent RB, Wood CR (2006) Therapeutic peptides: technological advances driving peptides into development. Curr Opin Biotechnol 17: 638-642.

50. Sato AK, Viswanathan M, Kent RB, Wood CR (2006) Therapeutic peptides: technological advances driving peptides into development. Curr Opin Biotechnol 17: 638-642.

51. Woodley JF (1994) Enzymatic barriers for GI peptide and protein delivery. Crit Rev Ther Drug Carrier Syst 11: 61-95.

52. Witt KA, Gillespie TJ, Huber JD, Egleton RD, Davis TP (2001) Peptide drug modifications to enhance bioavailability and blood-brain barrier permeability. Peptides 22: 2329-2343.

53. Bray BL (2003) Large-scale manufacture of peptide therapeutics by chemical synthesis. Nat Rev Drug Discov 2: 587-593.

54. Angelini A, Cendron L, Chen S, Touati J, Winter G, et al. (2012) Bicyclic peptide inhibitor reveals large contact interface with a protease target. ACS Chem Biol 7: 817-821.

Advances in Protein Chemistry

Chapter: New Protein Purification Approaches

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published November, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

New Protein Purification Approaches

Waseem Ahmad Siddiqui*

Department of Biochemistry, Jamia Hamdard (Hamdard University), Hamdard Nagar, New Delhi, India

***Corresponding author:** Waseem Ahmad Siddiqui, Faculty of Science, Department of Biochemistry, Jamia Hamdard (Hamdard University), Hamdard Nagar, New Delhi, 110062, India, E-mail: was.sid@gmail.com, wasiddiqui@jamiahamdard.ac.in

Abstract

The characterization of proteins and elucidation of its exact biological functions has always been a fascinating task for the researchers. Whether it is the study of proteins or obtaining it for commercial purpose, the development of techniques and methods for protein purification has remained an essential pre-requisite for all these advancements. Based on the properties of protein and nature of requirement, protein purification varies from simple one-step precipitation procedures to large scale validated production processes. Often more than one purification step is necessary to reach the desired purity. For successful and efficient protein purification, it is necessary to select the most appropriate techniques, optimize their performance to suit the requirements and utilize them in an appropriate way to maximize yield and minimize the number of steps required. Although, many different techniques are available, most purification schemes involve some form of chromatography. As a result chromatography has become an essential and powerful tool in every laboratory where protein purification is needed. The development of recombinant DNA techniques has revolutionized the production of proteins in large quantities. Recombinant proteins are often produced in forms which facilitate their subsequent chromatographic purification. However, this has not removed all challenges. Host contaminants are still present and problems related to solubility, structural integrity and biological activity can still exist. Although there may appear to be a great number of parameters to consider, with a few simple guidelines and application of the Three Phase Purification Strategy the process can be planned and performed simply and easily, with only a basic knowledge of the details of chromatography techniques. It gives an overview of the available methods and provides advice on how to avoid pitfalls in all operations, from initial sample preparation to final analysis of the purified protein. Today most of the purifications are carried out with affinity-tagged proteins. This approach greatly simplifies purification of many proteins. It is, however, somewhat common that the target protein is unstable under the conditions used for purification, or it may be difficult to obtain in sufficiently pure form. In such situations, more thorough purification protocol development may be needed to establish suitable conditions for purification of active protein.

Introduction

Protein purification has been developed in parallel with the discovery and further studies of proteins. Protein purification has been performed for more than 200 years when Antoine Fourcroy isolated some proteins from plants that had similar properties to albumin in 1789 (Table 1). However, until the beginning of the 20th century, the only available separation methods were filtration, precipitation, and crystallization. Table 1 very briefly enlists some of the milestones in the history of protein purification. The selection of milestones is a matter of individual perception and might well be a matter of argument.

Method	Year	Developed by
Protein Precipitation	1789	Antoine Fourcroy
Chromatography	1903	Tswett
Ultracentrifugation	1924	Svedberg
Moving boundary electrophoresis	1937	Tiselius
Ion exchange chromatography	1940s	The Manhattan Project
Two-phase partitioning	1955	Albertsson [1]
Size-exclusion chromatography	1955, 1959	Lindqvist & Storgards [2]
(Gel filtration)*		Porath and Flodin [3]
Sephadex™ (gel filtration medium)	1959	Pharmacia (now GE Healthcare)
Polyacrylamide gel electrophoresis	1959	Raymond and Weintraub [4]
SDS-PAGE	1967	Shapiro, et al. [5]
Affinity chromatography	1968	Cuatrecasas, et al. [6]
Two-dimensional chromatography	1975	O'Farrell [7]
Capillary electrophoresis	1981	Jorgenson and Lukacs [8]
Fast Protein Liquid Chromatography	1982	Pharmacia (now GE Healthcare)
*1955; Performed on Starch, 1959; Introduction of cross-linked dextran.		

Table 1: Some important advances in the history of protein purification.

During the 1950s and 1960s several new hydrophilic chromatography matrices were developed. In 1955 starch was used to separate proteins based on differences in size. Later, cross-linked dextran was developed which was more suitable for this purpose. During the 1960s, more materials for electrophoresis and chromatography where developed such as polyacrylamide, methacrylate, silica, and agarose. Rolf Axen introduced cyanogen bromide activation in 1967 that allowed ligand coupling to polysaccharides.

During the 1970s and 1980s a large number of chromatography media were developed, that lead to the revolution in protein

purification methodology and laid the foundation for the purification techniques that are in use today. In early 1980s, GE Healthcare (then Pharmacia) launched a completely integrated chromatography system called FPLC Fast Protein Liquid Chromatography (FPLC) system. FPLC has since become a useful tool for reproducible chromatographic purification of proteins. During that time, the emphasis was focused to purifications starting with natural sources, where extremely low concentration of target protein in the source organisms often made purification of even a few micrograms difficult and time consuming. Nowadays, Process-scale protein purification for large amounts of proteins used in laundry detergents or for enzymatic synthesis of complicated substances, as well as in biopharmaceuticals, has become common.

In 1970s, the development of recombinant DNA technology allowed modification and heterologous over expression of a selected protein. During the 1980s and 1990s affinity tagging of proteins became popular, and it allowed efficient affinity-based purification. The time of purification was substantially reduced since the same affinity tag could be used on many different proteins. Another important development in this area was during the Human Genome Project (1990 to 2003) was peptide mass fingerprinting, where Mass Spectrometry (MS) of peptide fragments of proteins combined with searches of databases with known protein sequences allowed identification of proteins present in a sample.

At present, protein purification is performed at scales ranging from micrograms and milligrams in research laboratories to tons in industry. Today, the purification approaches based on affinity tagging of the target protein have revolutionized protein purification. With its help, many proteins can be purified very easily and efficiently. However, it should be realized that even today, some proteins may be very challenging to purify in an active and stable form, these include unstable protein complexes, insoluble aggregates, integral membrane proteins, and proteins that are subjected to multiple post-translational modifications. The challenges in the area of protein purification that still persist make it not only worthwhile but also essential to gain a deep knowledge about protein purification so that the available methods can be selected and applied in an optimal way and can be improvised to suit the requirements. In addition, the doors always remain open for the development of new approaches to protein purification.

Protein Quantification, Yield and Purity

During the protein purification steps, it is always necessary to keep checking:

a. How much fold purification has been achieved till a given point? In other words, one has to look for how much of the contaminating protein have been removed.

b. What is the yield of the obtained protein? This gives an idea about the percentage of protein that has been retrieved out of the starting amount or activity.

To achieve these objectives, one needs to have:

a. A method for the estimation of total protein: It is required since the specific assay of target protein gives no information about the presence of contaminating proteins in the sample.

b. A specific assay method for the desired protein: This is meant for getting an idea of the amount of target protein and degree of purification achieved.

Protein estimation

There are several approaches to quantify the proteins. However, many of these, for instance, Radiolabelling, Edman degradation, RP-HPLC, Mass spectrometry, etc., are used less frequently or in special cases due to being more expensive, tedious and complex. In addition, there are other methods such as Kjeldahl method and Biuret assay that are relatively outdated and therefore, rarely find the use. In general, spectroscopic and colorimetric methods are most frequently used for estimation of total protein in a sample at a given stage. Basic principle of some of the commonly used spectroscopic and colorimetric assay methods are briefly mentioned below;

a. Biuret method: Molecules containing two or more peptide bonds form react with copper (II) under alkaline conditions to form a purple colored complex that absorbs maximally around the wavelength of 540 nm. The intensity of color is directly proportional to the concentration. The method I rarely used alone due to low sensitivity. However its variants, such as Lowry's method and Bicinchoninic acid method (discussed below), are commonly used.

b. Lowry's method: The method combines the Biuret test (i.e. the reaction of copper (II) with the peptide bonds under alkaline conditions), with the oxidation of aromatic amino acid residues in the protein. A purple-blue color is obtained by the reduction of phosphotungstic - phosphomolybdic acid in the Folin–Ciocalteu reagent to heteropolymolybdenum blue by oxidation of aromatic acids, catalyzed by copper (I), produced in the oxidation of peptide bonds. The absorbance is taken at 660 nm.

c. Bicinchoninic Acid (BCA) method: Like Lowry's method, BCA assay primarily relies on two reactions. First, the reduction of cupper (II) to copper (I) by the peptide bond occurs under alkaline conditions. Second, the chelation of each copper (I) ion by two molecules of BCA, forming a purple-colored product that strongly absorbs light at a wavelength of 562 nm. However, unlike Lowry's method, a single reagent containing all the components required for both steps, is sufficient.

d. Bradford's Dye-Binding method: The method involves the binding of Coomassie Brilliant Blue G-250 to proteins primarily through ionic interactions with basic amino acid residues (hydrophobic interactions are also believed to exist). The binding of the dye to protein shifts the absorption maximum of the dye from 465 to 595 nm, which is monitored. The assay is reproducible, single step and rapid (2 minutes incubation). The color remains stable for about one hour.

e. Absorption at 280 nm: Proteins show characteristic absorption at 280 nm. Aromatic amino acid residues Tryptophan, followed by Tyrosine are major contributors to absorption. Phenylalanine and Disulfide bonds also absorb at this wavelength but very weakly. The method is simple and does not lead to protein denaturation and sample loss. However, the method is error prone due to the interference by the contaminants, such as nucleic acids that absorb in similar range.

f. Absorption at 205 nm: Proteins absorb at 205 nm due to the absorption by peptide bonds. It may be useful when there is interference with absorption at 280 nm or there are very few or no aromatic amino acids in the protein. However, several buffer components and solvents absorb in this region and may lead to error.

g. Fluorimetric methods: The fluorescence of tryptophan (at 348 nm) is generally measured. The fluorescence intensity of the protein sample is used to calculate the concentration from a calibration curve obtained from the fluorescence emission of standard protein solution. Fluorescence may also be determined after derivatization of proteins by using fluorescent tags such as o-phthalaldehyde and fluorescamine.

Specific assay method

In order to follow the purification of a protein, it is necessary to have a method for specifically detecting and estimating the target protein in the sample. A specific assay method for the protein makes it possible to determine the yield, that is, how much of the desired protein has been purified, as well as the fold purification, which gives an idea of the degree of removal of contaminants. The later also requires a total protein estimation method. If the protein is colored (e.g. Cytochromes, Myoglobin) or fluorescent (e.g. Green fluorescent protein), the detection is simple and easy. However, these cases may be quite rare since most of the proteins are colorless. If the target protein is enzyme, an assay method may be developed that is based on monitoring the rate of formation of product or rate of utilization of the substrate. Such assays are usually colorimetric or spectrophotometric. For the non-enzyme proteins, the assay may be based on their biological activity or immunological property. While dealing with protein purification, calculation / estimation of the following are always a part of the protocol.

a. Units: Determined by using specific assay method, as discussed above, based on biological (e.g. catalysis) or physical (color, fluorescence) property of the target protein.

b. Total activity (Total units): This is a measure of how many units of protein are there in the sample at a given stage. It can be obtained by multiplying the activity (units per unit volume) in the sample by the total sample volume.

c. Specific activity: It is a way to find out the degree of purity (or degree of contamination for that matter) of the target protein. Dividing the activity/units by the amount (in mg) of total protein gives the value of specific activity. The higher this value, the higher is the purity.

d. Fold purification: It is calculated by dividing the specific activity of each fraction by the specific activity found in the initial sample with which the purification was started. The value changes with the type of protein and therefore, no minimum or maximum value can be fixed. However, this value, along with the percent yields (see below), indicates if a step was worthwhile or not. A step with poor fold purification with a low yield may be avoided, if possible, while the opposite (high fold purification and high yield) will be highly desirable.

e. Percent yield: It is the percentage of the total activity/units obtained after a given step of the total activity from the starting step. Final yield may be calculated by finding out the percent of protein activity/units obtained after the final step, compared to initial sample.

Modern Protein Purification Approaches

Efficiency, economy and sufficient degree of purity and quantity, are the common primary objectives of any protein purification procedure. It applies equally well to preparation of protein sample for biochemical characterization and to large scale production of any commercially important protein. It is, therefore, necessary to set objectives for purity, amount, retention of biological activity and economy for any process.

The following points must be taken into account for selecting or designing a purification process:

a. The nature and composition of the source material. It is important to differentiate between contaminants that need to be eliminated and those which can be tolerated. The information regarding the properties of the protein of interest and contaminants will allow faster and easier technique selection and avoid the inactivation or interference with the activity of the target protein.

b. The intended use and application of the final product, that is, whether the product is required for analytical or preparative use, what is the minimum percent purity required.

c. Safety issues, if any.

Apart from these, other factors can also influence the nature of objectives. Since there exists a great variation in the properties of proteins and the chemical nature and complexity of the sources of these proteins, there is no standard protocol for purification. For example, body fluids such as plasma is homogenous and requires simple handling after collection, on the other hand, tissue samples require extensive processing in order to obtain protein. Chances of loss of biological activity or structure may be of greater concern if harsher conditions are applied. Since many studies may require functionally active proteins, an utmost care must be exercised in these cases. Some of these strategies to minimize the losses may be:

a. Quick freezing of the samples using liquid nitrogen.

b. Use of stabilizing agents for the protein of interest.

c. Addition of inhibitors of the enzymes that may affect the protein of interest.

d. Consideration of time of exposure to extraction buffer or solvent as well as the solubilizing agent used.

Overall, if possible, a simple strategy with fewer steps is always helpful in minimizing the protein losses. Solubilization of protein is usually the preliminary step while dealing with tissue proteins. It is the process of breaking the interactions (ionic and non-ionic) that lead to aggregation and precipitation of protein that may lead to sample loss. The choice of buffer is always important as the protein solubility is influenced by both pH and ionic strength. In addition, the use of different buffers may lead to extraction of different sets of proteins form the same sample. Use of mild non-ionic (such as Titon-X-100) or ionic (such as CHAPS) detergents may sometimes be useful, especially in the solubilization of membrane proteins. Small amounts of reducing agents e.g. Dithiothreitol (DTT) and Beta-Mercapto Ethanol (BME) are usually added to maintain a reducing environment which may be essential for many proteins.

An overview of various purification techniques is given in the following sections. The description of techniques used in protein purification in the upcoming sections has been kept short and some minute details have been skipped in order to keep the chapter concise.

Fractionation of proteins in solution

Almost invariably, the target protein is obtained in aqueous solution together with several undesirable proteins impurities. The fractionation techniques may provide a way to remove at least some of the impurities while reducing the bulk of the sample. Although fractionation does only lead to a very partial purification, it generally precedes the high resolution techniques. In this way, a crude sample is partially purified and its volume is reduced to make it suitable to be processed further by more sophisticated methods. Depending upon the properties of the target protein, one of the following methods may be used for fractionation.

Fractionation by precipitation: This method utilizes the property of the differential solubility of proteins under different conditions as well as conditions which allow aggregation of proteins. The precipitation may be enforced by altering the buffer composition. The precipitated protein may be separated from other simply by low speed centrifugation. One of the following methods may be used to achieve this objective:

a. Salt fractionation: High ionic strength promotes protein precipitations. With increasing ionic strength, proteins begin to interact via hydrophobic patches on their surface, as proteins and salt compete for the residual water. This leads to formation of aggregates and the precipitation of proteins. The phenomenon is termed as "Salting out" of proteins. Different proteins would precipitate at different salt concentrations. Ammonium sulfate is used as the salt of choice, since it preserves protein activity and promotes precipitation at lower concentrations than other salts. Other salts may give low ionic strength (e.g. NaCl, KCl), have solubility problems (e.g. Na_2SO_4) or may cause damage to protein structure. Increasing amount of ammonium sulfate is added to the extract, with intermittent centrifugation steps. These stepwise precipitations are referred to as ammonium sulfate "cuts". The presence of target protein is checked in these cuts and an appropriate cut is determined where the major portion of target protein is concentrated. Pre-calculated tables are available to find out the amounts of solid ammonium sulfate to be added to a known volume in order to obtain the desired percentage saturation.

b. Isoelectric precipitation: Apart from ionic strength, the solubility of proteins also depends on pH. As the pH approaches the pI of a given protein the net charge on that protein gets closer to zero, and weak electrostatic attractive forces lead to aggregation and precipitation. Lowering ionic strength, or addition of solvent, promotes these interactions. Although, solubility is minimal at the isoelectric point, the increase in ionic strength increases the solubility even at the isoelectric point (This effect is called "salting in" of proteins).

c. Solvent precipitation: The mechanism of solvent fractionation is reduction of availability of water for protein solvation (i.e. a reduced water activity) which causes the precipitation of the protein. It tends to be more effective at pI which suggests that the mechanism may be similar to that in isoelectric precipitation. The precipitation may be achieved by adding water-soluble organic solvents such as ethanol or acetone. The concentration of solvent required depends on the type of protein and usually ranges between 5 to 60% (v/v). Since the denaturing effect of solvents increases with temperature, solvent fractionation is generally carried out around 0°C. The method has been extensively used in the past but nowadays only done in few specific cases. The main disadvantage of the method is the denaturation of proteins due to non-polar solvents. However, if the protein of interest is known to be stable in these conditions, the method may be worthwhile to use since it will remove many of unwanted proteins that are sensitive to solvents.

d. Selective denaturation: Contaminant proteins in isolated cases can be selectively denatured by heat, extremes of pH, and organic solvents. These treatments may be carried out if the protein of interest is stable towards one or more of these conditions. However, the treatment may result in chemical modification of the desired protein.

It is important to note that whole of the protein of interest may not come in a single fraction. Many times, some of the protein has to be sacrificed in order to get a better elimination of impurities.

Ultra filtration: Dialysis, that employs a semi permeable membrane made from cellulose, is frequently used for desalting purposes. The protein is retained in the dialysis bag while the salt moves out. If the membrane has larger pores that also allow small proteins to move out while retaining the larger ones, such membranes can be used for protein fractionation. Since dialysis speed is a function of molecular mass, the movement of proteins through dialysis tubing would be insignificant. Ultrafiltration, on the other hand, uses such membranes and employs pressure to force the sample through. Membranes ranging from 10kD to 100kD molecular weight cutoffs are commercially available. The method is not very efficient due to significant loss of sample, but may be useful in certain specific cases.

In order to increase the processing potential and the specificity of membrane filtration, affinity ultrafiltration concept was introduced. Affinity ultrafiltration follows the principle that

a. When the target protein and impurities are free in a solution pass via the ultrafiltration membrane.

b. Whereas, when a macro-ligand is bound to the target protein, the resultant is a membrane-retained complex.

After the target protein binds to the macroligand, the impurities are washed away. Dissociation of the target protein from macroligand can be performed by providing suitable conditions.

Solid Phase Extraction (SPE)

Solid Phase Extraction (SPE) is method that uses a solid phase and a liquid phase to isolate one type of analyte from others based on physic-chemical properties. It is generally used to concentrate and/or clean up a sample before using a chromatographic or other analytical method. Solid-phase extractions use the same type of stationary phases as are used in liquid chromatography (discussed in later sections).

SPE uses the affinity of solutes dissolved or suspended in a liquid phase (the mobile phase) for a solid phase (the stationary phase)

through which the sample is passed to separate a mixture into desired components and impurities. The result is that either the analytes of interest or the impurities in the sample are retained on the stationary phase. The other component passes through the mobile phase. Depending upon the situation, the phase that contains the desired analyte is processed for further purification.

Briefly, four types of sorbent formats exist [9]:

a. Cartridge: It is a small plastic or glass open-ended container filled with adsorptive particles of various types and adsorption characteristics. The cartridge type is still the most popular format

b. Disc: Disk is a variation of the extraction cartridge. They consist of a 0.5 mm thick membrane where the adsorbent is immobilized in a web of micro fibrils. The sorbent (on polymer or silica) is embedded in a web of PTFE or glass fibre.

c. SPE pipette tip: The solid phase sorbent is positioned inside a pipette tip, held in place by a screen and filter. The stationary phase is mixed with the sample.

d. 96-well SPE plate: Each of the 96 wells has a small 1 or 2 ml SPE column with 3-10 mg of packing material. The packing material in an SPE cartridge is placed between the bottom frit or membrane and the top frit.

Solid phase extraction formats are available with a variety of stationary phases, depending on the properties of protein to be isolated. Most stationary phases are based on silica that has been bonded to a specific functional group such as hydrocarbon chains (for reversed phase SPE), quaternary ammonium groups and sulfonic acid groups (for ion exchange SPE).

On the basis of the sorbent used, SPE may be of following types;

a. Normal phase SPE: In normal phase SPE, the stationary phase is usually composed of polar functionally bonded silica with short carbons chains. This stationary phase will adsorb polar molecules which can be collected with a more polar solvent.

b. Reversed phase SPE: The stationary phase in reversed phase SPE is derivatized with long hydrocarbon chains. Less polar compounds are retained due to hydrophobic interactions. The analyte can be eluted by using a solvent with a lower polarity.

c. Ion exchange SPE: It separates analytes on the basis of electrostatic interactions between the charged stationary phase and oppositely charged analytes in the sample. Ion exchange SPE may be carried out either with Cation exchange or with Anion exchange sorbents, depending upon the charge on desired analyte under given conditions. To elute the analyte, the stationary phase is washed with buffer that neutralizes the charge of either the analyte or the stationary phase. Once the charge is neutralized, the electrostatic interactions cease to exist and the analyte is eluted.

A typical solid phase extraction involves four basic steps (Figure 1). First, the cartridge is equilibrated with a buffer of the same composition as the sample. The sample is then added to the cartridge. As the sample passes through the stationary phase, the desired proteins in the sample will interact and retain on the sorbent. The washing will then remove the impurities that will pass through the cartridge. Then, the protein is eluted with appropriate buffer.

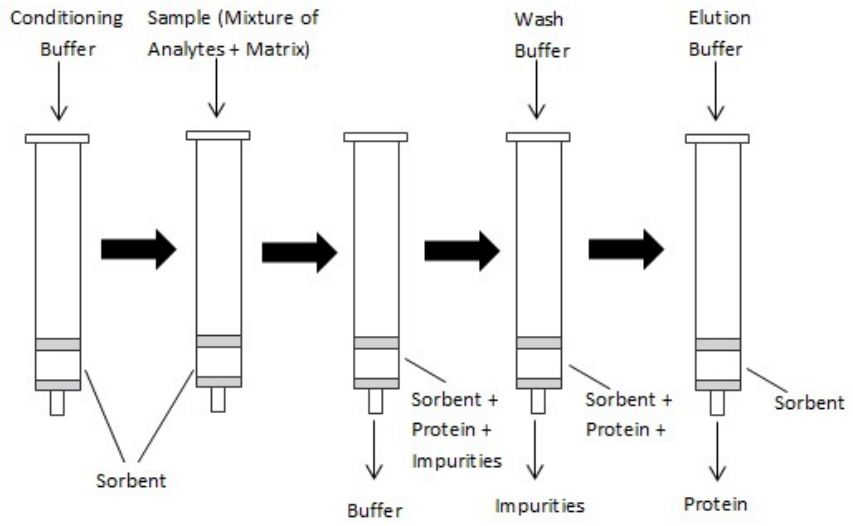


Figure 1: Steps in Solid phase extraction of proteins.

SPE offers several advantages over traditional liquid-liquid extraction techniques, such as higher recovery and concentration of analyte, range of samples that can be used and ease of automation [9].

Chromatographic techniques

Chromatography is a separation process based on distribution between two phases, a solid or liquid stationary phase and a liquid or gas mobile phase. The sample is propelled by fluid mobile phase which percolates the stationary phase. Different chromatographic processes based on characteristic principles may be used for the separation of a wide variety of substances including proteins. A given chromatographic process may be run both in low and high pressure systems, although, the basic principle remains same.

Most of the protein purification procedures involve one or more of the chromatographic techniques. Based on the properties and available information about the target protein, one can choose from various available chromatography methods.

Affinity chromatography (AC): The principle of AC is the separation of proteins on the basis of a reversible interaction between the target protein and a specific ligand bound to the matrix. This protein-ligand interaction can be biological in nature or biospecific (such as Antigen-Antibody binding, Immunoglobulin-Protein A binding, Receptor-Hormone binding, Enzyme-Substrate binding, etc.) or non-biospecific (such as protein-dye binding, Histidine containing proteins-Metal ions binding). Due to its specificity, it provides a high resolution, and moderate to high capacity. The sample is applied under conditions favoring selective protein-ligand binding. Unbound material is washed out, and by applying the conditions that favor weakening of protein-ligand interaction, elution is then achieved. Elution may either be carried out specifically, using a ligand that competes with the bound ligand for binding to the target protein, or nonspecifically, by simply altering the conditions of the medium such as pH, Salt concentration, or polarity that favor the dissociation of the protein from bound ligand (Figure 2).

Due to its high specificity, resolution and capacity, AC can sometimes be used for single-step purification especially when minor impurities can be tolerated. However, it is more commonly used as the first purification step.

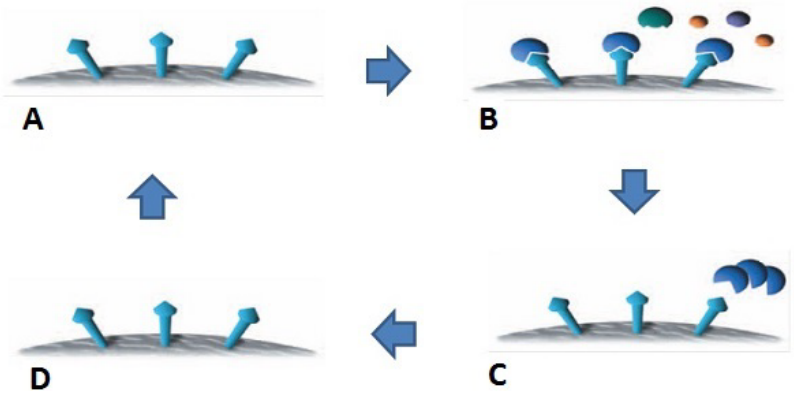


Figure 2: Scheme of Affinity chromatography
A. Equilibration of affinity medium is equilibrated with binding buffer.
B. Application of sample under conditions favoring reversible binding of the target molecule(s) to the ligand. The unbound material is washed out.
C. Elution of the target protein by changing conditions to favor dissociation of the target protein from the ligand of the bound molecules. It can be performed specifically (using a competitive ligand), or non-specifically (by changing the pH, ionic strength or polarity).
D. Re-equilibration of affinity medium with binding buffer [Adapted from 'Affinity Chromatography, Principles and Methods', Amersham Biosciences].

To increase the efficiency of binding of ligand to the target molecule, a spacer arm is introduced between the ligand and the matrix. This helps in avoiding the problem of steric hindrance that may occur if the target molecule binds to the ligand directly attached to the bead. Spacers are more important if the ligand is small in size. The length of a spacer arm ranges between 6-10 carbon atoms or a similar distance. A spacer must not affect the sample or the ligand or the binding between the two. Immobilization of a ligand on the matrix involves three steps.

a. Activated of matrix to make it reactive toward the ligand. The matrix should be activated in a way that allows covalent attachment of a ligand with ease.

b. Covalent coupling of the ligand to the matrix.

c. The unreacted groups are blocked by using an excess of a suitable substance that is not supposed to interfere with the protein ligand binding (ethanolamine is one of the commonly used substance). This is to make sure that the binding occurs only at desired sites.

AC simplifies the isolation process by using pre-existing, highly specific ligand-binding properties. As a result, AC is capable of giving very high purification, even from complex mixtures, in a single step. It retains its utility even when the target protein is in small amount in the mixture. Another objective that can be achieved through AC is the separation of biologically active form of the molecule from its functionally altered or its denatured form, since it is primarily based on functional properties. In addition, AC may also be employed for removal of specific contaminants.

Nowadays, an increasing number of laboratory-scale purifications are performed with affinity-tagged proteins. A variety of different affinity tags, chromatography media and pre-packed columns are offered to choose optimal conditions for different proteins and purification procedures. The most common examples are the purification of Histidine-tagged proteins using Immobilized Metal Affinity Chromatography (discussed in later section) or Glutathione S-transferase-tagged proteins using immobilized glutathione.

Immobilized metal ion affinity chromatography (IMAC): Immobilized Metal Affinity Chromatography (IMAC) is based on the formation of coordinate bonds between metal ions and some amino acids residues exposed on the surface of proteins (mainly histidine, but also tryptophan and cysteine). Since it is not a biospecific interaction in true sense, IMAC may be considered as a pseudo-affinity technique. The very principle of IMAC was first introduced by Porath et.al. [10]. But it has gained a deeper interest only recently due to the development of better and stable adsorbents and development of genetic engineering techniques that allow the introduction of histidine tags to proteins. The histidine tags are attached to the N- or C-terminus of recombinant proteins. Different histidine tags, from very short, such as His-Trp, to long ones, containing up to ten histidine residues, have been used. Currently, His6 tag (consisting of six consecutive histidine residues), is one of the most frequently used tags purification of recombinant proteins.

The affinity of a protein for a metal depends mainly on the metal ion involved in coordination. Divalent ions of transition metals, such as, Zn^{2+}, Ni^{2+}, Co^{2+}, Cu^{2+} and Fe^{2+}, are commonly used. These metal ions are immobilized on the matrix by using a suitable spacer arm plus a simple chelator which attached to the matrix . Some commonly used chelators are Imino Diacetate (IDA), Nitrotriacetic Acid (NTA) and Tris(carboxymethyl) Ethylene Diamine (TED).

As mentioned earlier, the interaction used in IMAC depends mainly on the formation of coordinate bonds between electron deficient metal ions and the electron donor groups on the protein surface. Side chains of certain amino acids are differentially suitable for binding. Histidine exhibits the strongest interaction, due to the presence of imidazole ring that readily forms coordination bonds with metal ions. To a lesser extent, sulfhydryl group of Cysteine, and aromatic side groups of Tryptophan, Tyrosine and Phenylalanine can also contribute to binding. However, the actual retention of protein is primarily based on the presence of histidine residues.

Since most of the proteins contain these amino acids in varying amounts, all proteins should theoretically be capable of binding to immobilized metal. However, the actual strength of binding would depend on the number of these residues as well as their presence on the surface of protein.

Target protein can be eluted from an IMAC resin decreasing the pH to 4-5 which leads to the protonation of histidine's imidazole group and hence the dissociation of the protein. For proteins that are sensitive to acidic pH, competitive elution with imidazole at nearly neutral pH is recommended.

Although, IMAC cannot be considered as highly specific when compared to other techniques, it is atleast moderately specific. However, it offers advantages such as stability of ligand, high capacity, mild elution conditions, ease of regeneration and low cost makes it useful.

IMAC holds an advantage over biospecific affinity techniques due to its structure-independent interaction that makes it effective even under denaturing conditions. This is often desirable when proteins are expressed in *E. coli* in the form of inclusion bodies.

On the other hand, IMAC may not be the technique of choice for the production of therapeutic proteins, due to problems with reproducibility, and contamination by host cell proteins, toxins, viruses etc. Moreover, metal ion leaching from the sorbent can cause damage to the target proteins by metal-catalyzed reactions.

Immuno-Affinity chromatography and immunoprecipitation: Immuno-affinity chromatography, also known as immune-adsorption chromatography, is a specialized form of affinity chromatography. It is based on the highly specific antigen-antibody interactions, and utilizes an antibody (or antibody fragment) immobilized onto a solid support matrix as a ligand in a manner that it retains its binding capacity towards the antigen. The crude sample is passed through the column where the binding of protein with antibody occurs. The unbound material is then washed clear before the elution of the retained protein. The elution is performed by alterations of the buffer (mobile phase) conditions to weaken the antibody-antigen interaction. The technique may well be used for the separation of antibodies using immobilized antigen.

Immunoprecipitation is also based on antigen-antibody interactions and therefore it is, in principle, similar to Immuno-affinity chromatography. The affinity of Protein-A or Protein-G for the antibodies is usually exploited. In the pre-immobilized antibody approach (Figure 3 A), the target protein (antigen) is allowed to bind to its antibody that is pre-bound to Protein-A (or Protein-G) attached to a solid support (beads). Alternatively, free antibody approach may be used that where the attachment to Protein-A (or Protein-G) beads is carried out after the formation of antigen-antibody complex (Figure 3 B). In either case, the beads are then separated from the mixture by centrifugation and the antigen is retrieved. The technique may also be used for pulling down and isolating the whole protein complex if the target protein naturally exists in association with other known or unknown protein partners, and is often termed as Co-Immunoprecipitation in this case. However, care should be exercised in interpretation of results since sometimes some non-specifically associated proteins may also get pulled along with the target protein.

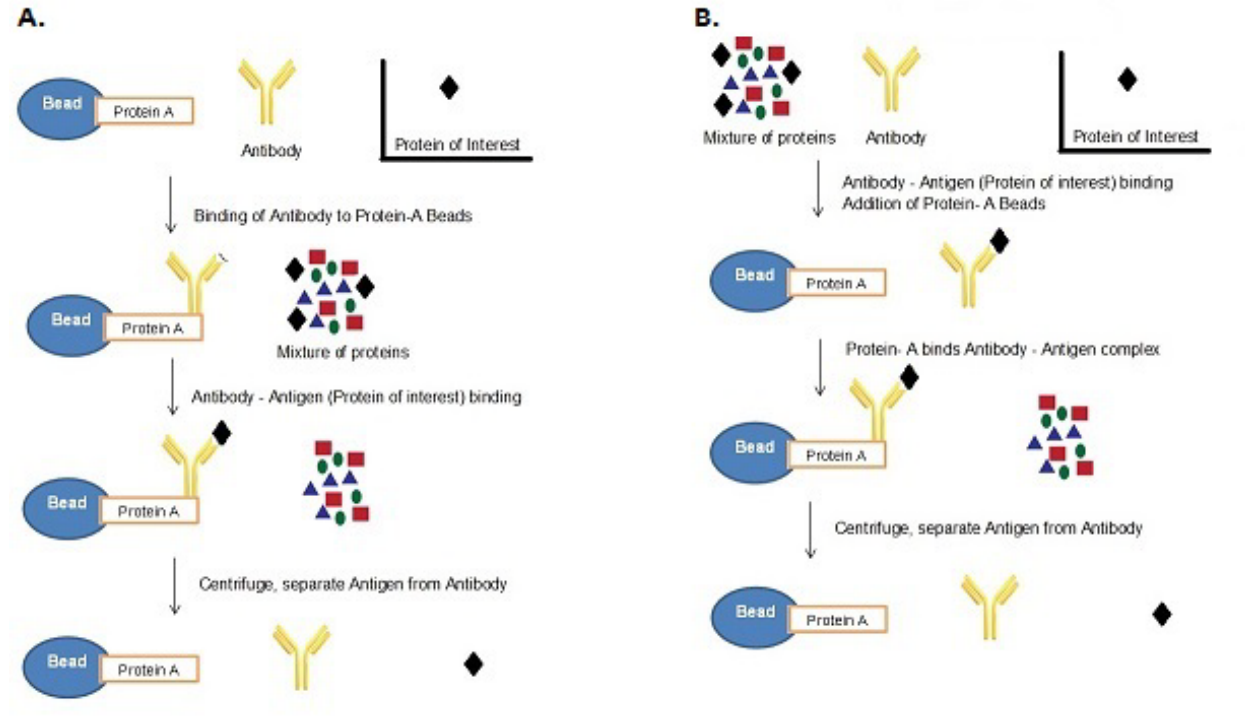


Figure 3: Steps in Immunoprecipitation. A. Pre-immobilized antibody approach B. Free antibody approach.

Ion exchange chromatography: Ion exchange chromatography (IEC) involves the reversible interactions between a charged protein and the chromatography medium with an opposite charge. The ion exchangers that were put in use initially were comprised of hydrophobic polymer matrices, heavily substituted with ionic groups. The low permeability of these matrices limited their use with larger

molecules such as proteins. Furthermore, they tend to denature the proteins due to their hydrophobic nature. In the early second half of the twentieth century, the introduction of hydrophilic materials of macro-porous structure made IEC a useful separation tool.

The net charge of proteins varies with the pH of the medium. A protein will possess a net negative pH above its isoelectric point (pI) and would bind an Anion exchanger more strongly, although it may still bind a cation exchanger, though weakly, as it would also possess some positively charged residues on the surface. An opposite phenomenon may be expected at pH lower than the pI value of protein. Figure 4 shows a diagrammatic representation of an ion exchanger. Some of the commonly used strong and weak ion exchangers are shown in Table 2. A strong ion exchanger binds across a larger pH range and more strongly as compared to a weak ion exchanger. Proteins bind as they are loaded onto a column in a buffer at desired pH. The ionic strength of the buffer is initially kept low. Elution can be performed by changing pH but more commonly by increasing salt concentration. The latter is preferable for proteins as pH alterations may lead to protein denaturation. The change in pH or ionic strength may be brought about in a stepwise manner or in form of a gradient. The most common salt is NaCl, but other salts such as sulfate, bromide or iodide of sodium or potassium may also be used.

The utility of IEC is evident from the fact that it may be used effective during any stage in a purification procedure. It is used as a first step, where capturing the target protein and reducing the bulk is desired. In middle or final stages of purification it can be efficiently used to achieve a better resolution of the sample. As the case with AC, IEC can also be used to bind and remove impurities, letting the target protein pass through. IEC can be repeated at different pH in order to achieve separation of several proteins. Alternatively, IEC with cation exchanger may be followed by anion exchanger to achieve multiple separations.

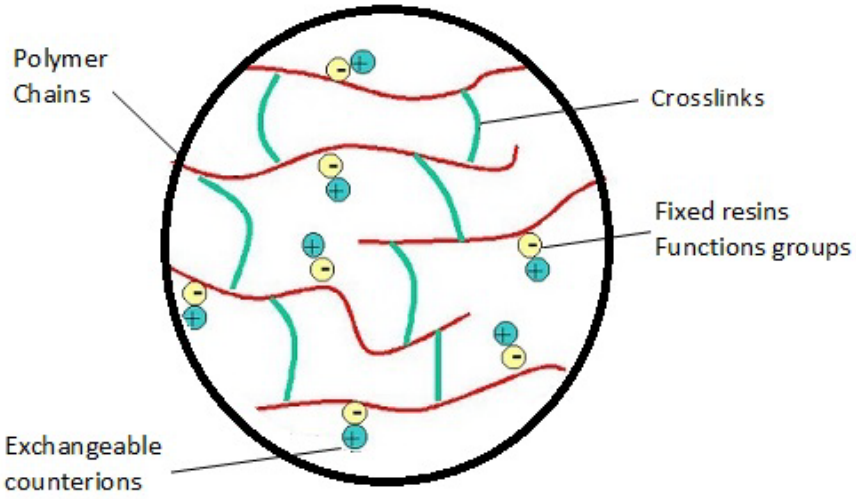


Figure 4: Diagrammatic representation of a Cation-exchange resin bead.

Exchanger Type	Ion exchange group	Buffer counter ions	pH range	Commercially available as
Strong cation	Sulfonic acid (SP)	Na^+, H^+, Li^+	4-13	SP Sepharose® SP Sephadex®
Weak cation	Carboxylic acid (CM)	Na^+, H^+, Li^+	6-10	CM Sepharose® CM Sephadex®
Strong anion	Quaternary amine (Q)	Cl^-, CH_3COO^-, $HCOO_3^-$, SO_4^{2-}	2-12	Q Sepharose® QAE Sephadex®
Weak anion	Primary/ Secondary/ Tertiary amine (DEAE)	Cl^-, CH_3COO^-, $HCOO_3^-$, SO_4^{2-}	2-9	DEAE Sepharose® DEAE Cellulose

Table 2: Some commonly used ion exchangers.

It can be concluded from the above discussion that IEC is amongst the highly efficient, powerful and frequently used protein purification techniques available not only for the separation of proteins, but also nucleic acids and other charged biomolecules. The technique possesses an advantage of simplicity and economy as the ion exchange resins may be used multiple times. Another advantage of IEC is the fact that the elution normally takes place under mild conditions, which is desirable to maintain the native conformation of the protein during the purification process. In addition, the non-specific interactions of non-ionic nature with proteins are low. However, a limited selectivity, which may lead to lower efficiency, may be considered as its main demerit.

Size exclusion chromatography: Size Exclusion Chromatography (SEC) allows separation of substances with differences in molecular size, under mild conditions. SEC is also referred to as Gel filtration chromatography, Gel permeation chromatography or Molecular sieve chromatography. SEC can be used for protein purification as well as removal of significantly small impurities, such as salts, from the proteins.

Molecular sieve properties of a variety of materials are utilized for separation in SEC. The matrices consist of beads with a range of pore sizes. The separation depends on the varying ability of various proteins to enter all, some or none of the pores or channels in the beads (Figure 5). The small molecules have more channels that they can enter and equilibrate in the mobile phase inside and outside the beads. This leads to a longer retention time and larger elution volumes. Very large molecules are excluded from the channels, and pass quickly between the beads and come out in minimum elution volume. The intermediate sized molecules take intermediate time for elution.

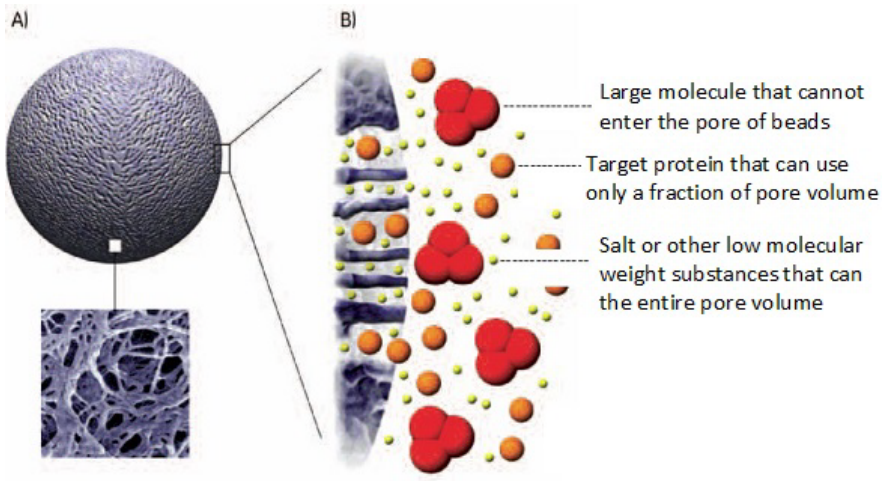


Figure 5: A. Diagram of a section of matrix bead. B. Diagrammatic representation of interaction of different sized molecules with the matrix [Adapted from 'Strategies for Protein Purification' Handbook by GE healthcare].

The matrices used in SEC are commonly composed of naturally occurring polymers such as dextran and agarose but synthetic polymers such as polyacrylamide are also available. Cross-linking of these polymers forms the three-dimensional network of the gels. Since the degree of cross-linking determines the pore size, altering the degree of cross-linking gives different pore sizes. Sephadex (composed of dextran) was the first commercial medium for GEC in which epichlorohydrin was used as a cross-linking agent. Agarose spontaneously forms gels without any cross-linking agent being used. In agarose the hydrogen bonding between the chains allows the formation of three dimensional structure of the gel.

Although the separation of molecules in SEC is generally considered to be according to molecular weight, it is more accurate to state that it is due to differential partition in the space between inside and outside of the porous particles. The degree of diffusion is determined by the hydrodynamic volume, which is the volume created by the movement of the molecule in the liquid. Linear molecules tend to have larger hydrodynamic volumes than globular molecules, so a 50,000 MW nucleic acid will elute much earlier than a 50,000 MW globular protein.

In SEC, there is no concentration of the sample components; rather the dilution of the sample takes place during the elution of the sample. It is therefore, recommended that the sample volume that is loaded on the column must be kept small to avoid the broadening of sample zone which would otherwise cause a lesser resolution. Sample volumes between 1% and 5% of the total column volume are generally used. However, one may use higher sample volumes if the size difference between the target protein and the impurities is high.

The elution involves a single buffer without any gradient being employed (isocratic elution). A broad range of pH, salt concentration, and temperature can be used with a variety of additives such as protein stabilizers, detergents and denaturants. The buffer system that suits the sample type and maintains the target protein activity is used.

SEC is rarely used as the first purification step but it proves to be a useful method for purification of proteins that have already undergone one or more purification or enrichment steps that lead to concentration of sample and the removal of bulk impurities.

Although SEC gives the low resolution, low capacity and high degree of dilution of the sample when compared to the other forms of chromatography, still, it is commonly used due to its simplicity. The main advantage of SEC is mild conditions and non-interaction with the sample, enabling the conservation of biological activity. Moreover, since no interaction is involved, SEC proves useful for the separation of multimers of the same protein that are not easily distinguished by many other methods. Due to these reasons, SEC is preferably used in final steps when sample volumes have been reduced and the protein has been fairly purified.

Hydrophobic interaction chromatography: Hydrophobic Interaction Chromatography (HIC) involves the reversible interactions between hydrophobic groups on the protein surface and hydrophobic ligands attached to matrix. The separation is on the basis of the amount of exposed hydrophobic amino acid residues.

The concept HIC was first given by Tiselius in 1948, when he first reported that proteins movement is retarded in a buffer containing salts which he termed as salting out chromatography. Initially, the matrices used were of a mixed hydrophobic-ionic character, which were later replaced by charge-free hydrophobic adsorbents which established the hydrophobic character of the interactions. The main contribution in these developments was of Hjertén [11] who also proposed the name of the technique as Hydrophobic Interaction Chromatography.

Many different types of matrices can be used for preparing HIC adsorbents, but agarose has been the most extensively used. Silica and organic polymers have also been used for adaptation of the technique for High-Performance Liquid Chromatography (HPLC). Since the charge may interfere with the elution of protein due to ionic interactions, the adsorbents should be charge free. Due to their pure hydrophobic character of alkyl groups (such as hexyl, octyl, etc.), linear chain alkanes are most commonly used. Sometimes phenyl group containing ligands may also be employed as they interact more selectively with aromatic groups.

Overall, the hydrophobic interactions between proteins and ligand are stronger when there is;

» Higher number of exposed hydrophobic groups on the surface of protein.

» Higher ionic strength.

» Longer carbon chain of ligand.

This is due to greater hydrophobicity of longer carbon chains. A chain of 4 to 10 carbon atoms is suitable in most of the cases.

However, for proteins with poor solubility in high salt buffers, such as integral membrane proteins, longer chain (12 to 18 carbons) adsorbents are generally recommended.

» Higher concentration of immobilized ligand in the matrix.

Although, the degree of interactions would increase with increasing concentrations, very high density of ligand is usually not recommended as the elution becomes difficult due to multi-point interactions.

While loading the sample, the conditions are set in a manner that promotes stronger binding of protein with the ligand (such as high ionic strength buffer), while elution may be performed by reversing the conditions (Figure 6). In general, the separations by HIC often use nearly opposite conditions to those of IEC. As far as the buffer system is concerned, HIC requires the presence of certain salt ions, which preferentially remove ordered water molecules surrounding the hydrophobic patches on proteins and thereby promote hydrophobic interactions. The elution is usually carried out by using very low salt concentration that weakens the hydrophobic interactions, although other parameters such as alteration in pH or polarity of the medium may also be employed. Cations and anions can be arranged in a series, called the Hofmeister series, starting with those that highly favor the interactions (also known as anti-chaotropes) to those that will reduce hydrophobic forces. Following is the list of some of the cations and anions in the above-mentioned order.

Cations; $NH_4^+ > Rb^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Ca^{2+} > Ba^{2+}$

Anions; $PO_4^{3-} > SO_4^{2-} > CH_3COO^- > Cl^- > Br^- > NO_3^- > ClO_4^- > I^- > SCN^-$

Most commonly used salts are ammonium sulfate ($(NH_4)_2SO_4$) and sodium sulfate (Na_2SO_4) as they are also known to stabilize protein structure. However, ammonium sulfate should be used at pH below 8 since it tends to decompose at higher pH. In addition, the salt concentrations used must always be below the concentration that salts out or precipitates any protein to avoid protein precipitation within the column.

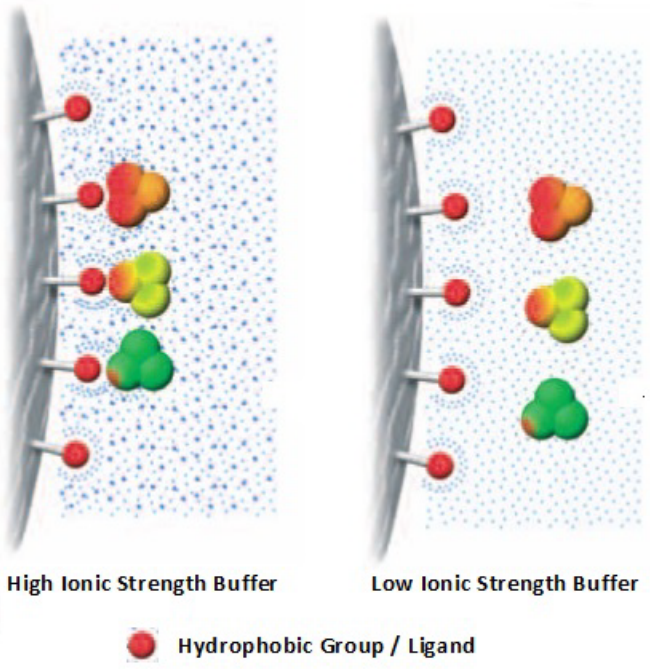


Figure 6: A. Figure representing the interaction of hydrophobic groups on the surface of proteins with the hydrophobic ligand immobilized on the matrix, which is facilitated in the presence high salt concentration. B. The dissociation of the proteins from the ligand at low ionic strength. [Adapted from 'Hydrophobic Interaction and Reversed Phase Chromatography, Principles and Methods', a handbook by GE healthcare].

The pH of the buffers has a decisive influence on the adsorption of proteins. However, many proteins may be denatured at very high or low pH values, as well as some supports may not be stable under certain pH conditions (for instance, silica is unstable at high pH), thus testing different pH values for adsorption is always desirable. The temperature dependence of elution or retention in HIC is quite complex, although, lower temperature facilitates elution and an increase in temperature would usually enhance protein retention. In spite of this, the temperature should be kept low keeping in view, the stability of proteins.

To conclude, it may be said that overall, HIC is mild method and the damage (structural or otherwise) to the biomolecules is minimal, mainly due to the stabilizing effect of salts and relatively weaker interactions of protein with the matrix. The recoveries are usually reasonably high.

Reversed phase chromatography: Reverse phase liquid chromatography (RPC) is the separation of molecules based upon their interaction with a hydrophobic matrix which is largely based on their polarity. The name "reversed phase" is derived from the opposite technique of "normal phase" chromatography which involves the separation of molecules based upon their interaction with a polar matrix (such as silica beads without any hydrophobic groups attached) in the presence of a low-polarity solvent. Though the technique was originally developed for the separation of relatively small organic molecules was later applied for the purification of proteins also, due to its high resolving power. The mechanisms of separation for small organic molecules and proteins, however, are quite distinct.

RPC is quite closely related to HIC (discussed in the previous section) since both are based upon interactions between hydrophobic patches on protein surface and hydrophobic ligands linked to the adsorbent. Thus, the basic molecular interactions are very similar in

both cases. Furthermore, in both cases, the sample is applied in the aqueous mobile phase. However, the adsorbents used for RPC are significantly more highly substituted with hydrophobic ligands than those used for HIC, due to which, the hydrophobic interaction in RPC is much stronger and can lead to adsorption of proteins even in absence of salt. However, these strong interactions usually require the use of low polarity medium and other additives to elute the protein, which may have a denaturing effect on the protein.

Porous silica beads are most commonly used matrix due to their mechanical strength, chemical stability in the solvents used for RPC, and Si-OH groups that can be modified to attach the hydrophobic ligand. However, this coupling is chemically unstable at pH values above 8. The free Si-OH groups released due to high pH may deprotonate and interact with counter-ions on protein surface leading to mixed (both ionic and hydrophobic) type of interaction of the matrix with the proteins. This may lead to abnormal retention and decreased resolution. Since residual silanol groups are believed to pose problems in separation, the silica particles should be completely covered hydrocarbon chains.

Shorter carbon chains (C2-C8) are used for protein separations, to avoid too strong interactions that require higher proportion of organic solvents for elution, which in turn, may cause damage to proteins. For smaller peptides, the use of longer aliphatic chains (C8-C18) is preferable.

Some synthetic organic polymers, such as polystyrene beads, are available and may be used in RPC. The surface being highly hydrophobic itself, does not require attachment of hydrophobic ligands. Furthermore, they are stable at high pH values but have lower mechanical strength as compared to silica.

The binding conditions in RPC are primarily aqueous mobile phase and the protein binds strongly to the matrix due to high hydrophobicity of the latter. This strong binding makes the use of organic solvents essential for elution. These solvents lower the polarity and of the mobile phase, facilitating the elution. While the binding of protein to the matrix is strong under aqueous condition, their desorption would only occur within very narrow range of concentration of organic solvent (also called organic modifier). Since most of the biological samples contain a mixture of different biomolecules with varying binding affinities, a gradient elution with increasing concentration of the organic solvent (that is, decreasing polarity of the medium) is generally performed.

Although, a wide range of organic solvents can be used, only a few are routinely used with acetonitrile and methanol being most frequently employed. Isopropanol, because of its strong eluting properties, may be used for elution but its high viscosity, which results higher column backpressures, limits its use. UV transparency of the mobile phase is a crucial property for RPC, since elution is usually monitored using UV detectors. Peptide bonds only absorb at low wavelengths in the ultra-violet spectrum where many organic solvents may also absorb. The presence of such solvent will cause a high background absorbance and may interfere with the interpretation of the results.

The strong binding and relatively harsh elution conditions in RPC may cause denaturation of proteins. Due to this reason, the method can be chosen for protein purification only under certain circumstances. Some of these are;

» When the protein can tolerate these conditions or may revert back to original confirmation afterwards (this may be true for some smaller proteins and oligopeptides).

» When even obtaining inactive or denatured protein may serve the purpose (e.g. for immunization, sequencing, etc.).

Since, a substantial number of cases do not fall in the above mentioned categories, therefore, RPC is mostly used for analytical studies of proteins where obtaining the biologically active form or native structure is not essential. However, if applicable, it proves to be a good separation technique with reasonably high resolution.

Chromatofocusing: Chromatofocusing (CF) separates proteins according to differences in their isoelectric point (pI). It is capable of resolving very small differences in pI (as low as 0.02 pH units) and is useful in separation of proteins that are quite similar to each other. Due to its low capacity it may only be used with partially purified samples. The buffer is consists of a large number of commercially available buffering substances (as Polybuffer™) and the medium is weak anion exchanger (due to the presence of amines). The interaction of buffer and chromatography medium leads to the generation of pH gradient. Proteins with different pI migrate at different rates in the column through the pH gradient, gradually concentrating in narrow bands and are eluted. The upper and lower limits of the pH gradient are defined by the pH of the start buffer and the elution buffer, respectively.

The medium is equilibrated with start buffer (any standard buffer with low ionic strength) at a pH slightly higher than the maximum pH desired. An elution buffer (Polybuffer, adjusted to the lowest pH required) is then passed through the column. The flow of this low pH buffers through the column lowers the pH and generates a descending pH gradient. The sample is then applied in the start buffer. Sample proteins which are at a pH above their pI (and hence are negatively charged) bind near the top of the column. Proteins that are at a pH below their pI start migrating downwards with the buffer flow and only bind when they reach a zone where the pH is above their pI. As the pH continues to decrease, a protein that that is subjected to pH below its pI becomes positively charged, and is repelled by the positively charged amine groups of the medium. This leads to migration of protein down the column at a velocity higher than the movement of pH gradient. Eventually, all the proteins are eluted from the column at a pH near their pI. The protein with the highest pI is eluted first and the protein with the lowest pI is eluted last. The proteins that easily precipitate at their pI should be avoided in CF.

CF provides high-resolution and it is equally good for analytical and in preparative purification where high purity is desired.

High performance liquid chromatography: High-performance liquid chromatography (sometimes referred to as high-pressure liquid chromatography), HPLC, is a chromatographic technique used to separate a mixture of compounds for various analytical and preparative purposes. HPLC finds enormous number of uses including those in medicine, research and industry.

HPLC is distinguished from traditional ("low pressure") liquid chromatography it operates at significantly higher pressures (50-350 bar), while ordinary liquid chromatography typically relies on the force of gravity and operates at normal atmospheric pressure. Also HPLC columns are made with very fine sorbent particles (2- 5 micrometer in average particle size) that can withstand high pressure generated during the separation.

The schematic of an HPLC instrument typically includes a sampler, pumps, and a detector (Figure 7). The sampler brings the sample mixture into the mobile phase stream which carries it into the column. The pumps deliver the desired flow and composition of the mobile phase through the column. The detector generates a signal proportional to the amount of sample component emerging from the column, hence allowing for quantitative analysis of the sample components. Some models of mechanical pumps in a HPLC instrument can mix multiple solvents together in ratios changing in time, generating a composition gradient in the mobile phase.

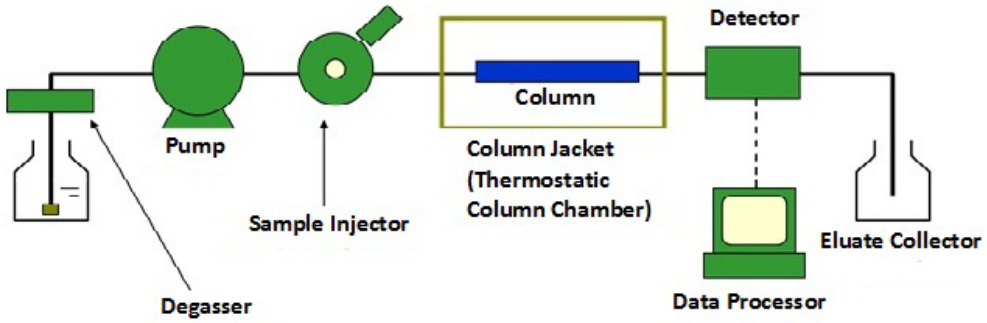


Figure 7: Flow chart diagram representing the components of HPLC system.

Though the principle of any chromatographic procedure remains same whether in normal or in high pressure mode, HPLC offers several distinct advantages over the normal pressure chromatography such as;

» High separation capacity, enabling the batch analysis of multiple components.
» Superior quantitative capability and reproducibility.
» Moderate analytical conditions.
» Generally high sensitivity.
» Low sample consumption.
» Easy preparative separation and purification of samples.

Fast protein liquid chromatography: Like HPLC, Fast Protein Liquid Chromatography (FPLC) is also an automated approach using chromatographic techniques. In HPLC, extremely high pressures and stainless steel columns that get heated may lead to protein denaturation. FPLC is developed to overcome these problems while using the similar principle. It is, therefore, also referred to as Protein-friendly HPLC.

The overall instrumentation of FPLC is similar as HPLC, but there are some basic differences in hardware as well as operating conditions. Some of these differences are mentioned below;

» In HPLC, high quality stainless steel is used whereas in FPLC, high quality plastics or glass is used (For temperature control).
» In HPLC, pressure of up to 550 bar is used, while much lower pressures are employed in FPLC (For the safety of column).
» In HPLC, standard analytical column (4-5 mm diameter, 10-30 cm length) is mostly used while in FPLC, microbore column (1-2 mm diameter, 10-25 cm length) is widely used.
» HPLC technique may not suit the separation of certain proteins, but FPLC is very reliable in separating & purifying proteins.

Due to the discussed reasons, FPLC finds enormous utility in various processes such as;

» Separation of thermally labile compounds.
» Qualitative and quantitative analysis of proteins.
» Quality assurance in pharmaceutical industry.

Like HPLC, this system consists of a programmable controller for developing and controlling automatic separation procedures, one or more pumps for liquid delivery, a mixer to ensure accurate and reproducible elution gradients, valves for sample injection and flow path control, one or more monitors for measuring chromatographic profile, a recorder for documenting chromatographic profile, a fraction collector and a chromatography rack for mounting the component in a compact laboratory bench top arrangement.

Several automated protocols have been developed recently to streamline the protein purification process. Each of these protocols is designed around an affinity chromatographic purification, either in a one-step, high-throughput approach or with affinity chromatography as the first step in preset dual-column protocols [12].

Electrophoretic techniques

The term electrophoresis refers to the movement of charged molecules in response to an electric field, resulting in their separation. Proteins, being charged molecules, also move in electric field towards oppositely charged electrode. The rate of migration is determined by the characteristics of proteins (size, shape, and charge) as well as the electrophoresis system (the electric field applied, the temperature, the pH, type of ions, and concentration of the buffer) [13]. Proteins exhibit a large range of shapes and sizes and possess varying degree of net charge. Due to these properties, every protein is expected to have a unique migration rate, a property which is used in their separation also. Protein electrophoresis is generally performed in gel-based media; however, free flowing systems for different types of electrophoresis have also been developed [14]. Protein electrophoresis can be used for a variety of applications such as protein purification, assessing the purity of obtained protein, studies of the regulation of protein expression, or determining protein mass, pI and biological activity.

Different electrophoretic methods help in achieving different objectives. Although, variety of gels may be used for electrophoresis,

the most frequently used are Agarose (more commonly for nucleic acids due to large pore size) and Polyacrylamide gels, which are more frequently used for proteins.

Polyacrylamide Gel Electrophoresis: Polyacrylamide gels (Figure 8) are formed by the polymerization of acrylamide (CH2CHCONH2), in the presence of free radical generator such as ammonium persulfate or riboflavin. The linear chains of polyacrylamide are cross-linked with the help of N, N methylene bisacrylamide ([CH2CHCONH]2CH2). Gels of different pore sizes may be formed by increasing or decreasing the percentage of monomers in the polymerization mixture. When the electric current is applied, proteins while moving towards opposite electrode experience a sieve effect due to the pores of the gel. This allows smaller proteins to move faster than the larger ones. For protein separation, virtually all methods use polyacrylamide that covers a size range of 5–250 kD. In PAGE, the gel is mounted between two buffer chambers, current flows through the gel. The gels may be run both in vertical and horizontal modes. The former is more common for polyacrylamide gels (Figure 9A).

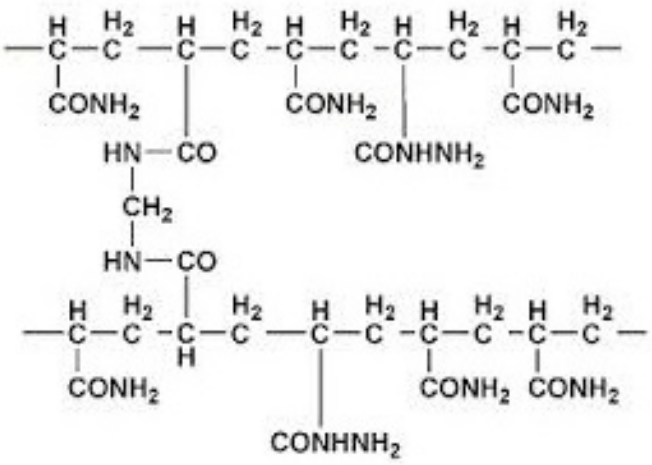


Figure 8: Structure of polyacrylamide.

Two types of buffer systems can be used:

» Continuous buffer systems; which use the same buffer (at constant pH) in the gel, sample, and the reservoirs. Continuous systems are not suitable for protein separations, and are used with nucleic acids.

» Discontinuous buffer systems; which use different buffer for reservoirs, the gel (sometimes two gels; a large pore stacking gel and a resolving gel) and the sample. These are generally employed with proteins.

Following electrophoresis, the gel may be stained (most commonly with Coomassie Brilliant Blue R-250 or silver stain), allowing visualization of the separated proteins, or processed further (e.g. Western blot). After staining, different proteins will appear as distinct bands within the gel (Figure 9B).

The following sections take up some of the polyacrylamide gel electrophoresis techniques that are utilized for purification of proteins.

Native Polyacrylamide Gel Electrophoresis: In Native Polyacrylamide Gel Electrophoresis (Native PAGE), proteins are prepared in a non-denaturing sample buffer, and electrophoresis is performed in the absence of any denaturing or reducing agents. Since the native charge and mass of the proteins is preserved, the movement is primarily according to its charge to mass ratio. However, additional factors (such as shape) that also influence the movement of proteins make the data from native PAGE quite difficult to interpret. Multi-subunit proteins may also be separated in their native form as the conditions are mild and non-denaturing. However, their movement is even more unpredictable. Depending upon charge, protein may move towards either of the electrodes. Native PAGE cannot be used for determination of molecular weight. Nonetheless, it does allow separation of proteins in their active state and can separate proteins of the same molecular weight on the basis of charge differences. Overall, it is a low resolution technique for purification but still a useful one due to discussed reasons.

PAGE may also be used as a preparative technique for the purification of proteins. For example, quantitative preparative native continuous polyacrylamide gel electrophoresis (QPNC-PAGE) is a method for separating native metalloproteins in complex biological matrices.

Sodium Dodecyl Sulfate: Polyacrylamide Gel Electrophoresis: Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE) was developed by Laemmli [15] in order to overcome the limitations of native PAGE. The detergent sodium dodecyl sulfate was incorporated in the discontinuous polyacrylamide gel electrophoresis buffer system (hence the name SDS-PAGE). However, the loss of native structure and biological activity in SDS-PAGE limits its use in condition where the isolation of protein in native state is desired.

In the presence of SDS (that disrupts all the non-covalent interactions in the protein) and a reducing agents (such as 2-mercapto ethanol, which reduces and breaks disulfide linkages within a polypertide or between the subunits), they become fully denatured and different subunits dissociate from each other. The non-covalent but uniform binding of SDS to proteins leads:

» An overall negative charge on the proteins by masking the intrinsic charge of protein.

» A uniform binding of SDS (1.4g of SDS per 1g protein, or approximately, one SDS molecule per two amino acid residues) generates a similar charge-to-mass ratio for all proteins in a mixture.

» Denaturation of all proteins gives them a similar rod-like shape.

This allows the movement of all the proteins towards positive electrode at a rate that depends only on the molecular size of the protein. The smaller proteins experience a lower degree of sieving in the get and move faster as compared to larger proteins. In SDS-PAGE, it is common to run proteins of known molecular weight (molecular weight markers) in a separate lane in the gel, in order to calibrate the gel and determine the approximate molecular mass of unknown proteins by comparing the distance travelled relative to the marker.

Apart from molecular weight determination, SDS-PAGE may also be effectively used for the checking the purity of a protein sample. It may also be a useful technique for the separation and purification of proteins, when obtaining a protein in its native form is not essential (such as for immunization. Automated SDS-PAGE systems are also available for preparative purposes (Figure 9C).

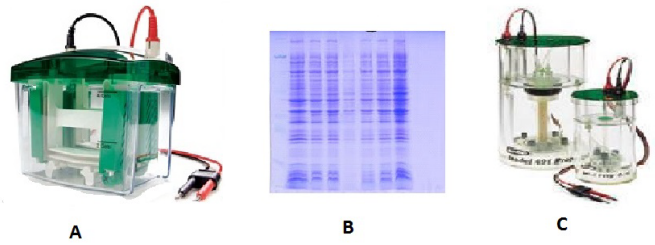


Figure 9: A. A verticl gel electrophoresis assembly (power supply not shown). B. A gel showing proteins bands after staining with Commassie Brilliant Blue R 250. C. A preparative SDS-PAGE assembly.

Isoelectric Focusing: The principle of separation in isoelectric focusing (IEF) is quite similar to chromatofocussing discussed in section 1.2.2.7. However, IEF combines the use electric current with a pH gradient in the gel in order to separate proteins according to their pI, which falls between pH 3 to pH11 for most of the proteins. Since the pI is a characteristic of a protein, IEF provides an excellent resolution. The movement of a protein through a pH gradient leads to a change in its net charge. Under electric field, the protein migrates to the pH value where its net charge becomes zero (that is its pI). The protein will stop at this position as due to a zero net charge it will not experience any pull under electric field. In other words, it is focused in the region of its pI. Similarly, other proteins in the mixture will also separate according to their pI values with a high degree of resolution.

A wide or a narrow range of pH may be chaosen for generating a gradient. A stable and continuous pH gradient between the between the electrodes is maintained by one of the following methods:

» Carrier ampholytes; which contain a heterogeneous mixture of small molecular weight compounds that carry multiple charges of both types, with closely spaced pI values. Chemically they are polyaminopolycarboxylate compounds. When the voltage is applied, the ampholytes align themselves in the gel according to their pI values and tend to buffer the pH in their proximity. This establishes a stable pH gradient.

» Immobilized pH gradients (IPG) strips; are formed by covalently linking the buffering groups to a polyacrylamide gel. A pH gradient of a desired range is generated by a gradient of different buffering groups.

IEF can be run under either native or denaturing conditions. Native IEF retains protein structure and enzymatic activity. Denaturing IEF is performed in the presence of high concentrations (usually 8M) of urea, which denatures the protein and dissociates the subunits, if linked non-covalently. Denaturing IEF often offers higher resolution and may give an insight into the type of interactions between subunits of multimeric proteins.

Two-Dimensional Gel Electrophoresis: Two-dimensional electrophoresis (2DGE) was first introduced by O'Farrell [16] and Klose [17] in 1975. The sequential application of two different electrophoresis techniques produces a multi-dimensional separation. In the most common 2DGE, protein samples is first separated in denaturing IEF in a tube gel or an Immobilized pH gradients gel strip to allow the separation on the basis of pI. This gel (or strip) is then equilibrated with SDS and placed on an SDS gel in the direction that is perpendicular to direction of separation in IEF. SDS-PAGE is then carried out in second dimension to separate the proteins according to their molecular weight. A very high-resolution is achieved in 2DGE due to the fact that it is highly improbable for two proteins to have same pI and molecular weight simultaneously. This method enables the separation of thousands of proteins (e.g. from cell lysate) in a single slab gel. The resulting spots can be visualized by gel staining.

3D Gel Electrophoresis: Due to a limited throughput, 2DGE may not be very suitable for comparative protein expression studies at large-scale. 3D gel electrophoresis (3DGE) may overcome this problem. It is relatively recent technique developed by Ventzki and Stegemann [18] that gives a much higher resolution than 2DGE and a larger number of samples can be simultaneously analyzed. Following conventional isoelectric focusing (IEF), up to 36 immobilized pH gradient (IPG) strips are arrayed on the top surface of a 3-D gel body, and the samples are transferred electrokinetically to the SDS gel. Since there may be unequal heat generation, special care is required here so that SDS-PAGE occurs under identical electrophoretic and thermal conditions, in order to avoid gel-to-gel variations. The proteins are labeled with fluorescent tag and the protein bands are resolved online by photodetection of laser-induced fluorescence (LIF). A digital camera placed below the gel captures the images in real time as the fluorescently labeled proteins pass through the laser-illuminated detection plane. Image processing software simultaneously processes these images, making results immediately accessible without further gel processing [19]. The separation patterns are analyzed using special software (available commercially e.g., Kapelan LabImage 1D or Decodon Delta2D), thereby improving the comparability, resolution and efficiency of the separation. Though still in the stages of development and improvisations, the method is set to find use in a wide range of applications in clinical diagnosis and pharmacology by

protein monitoring during disease development and screening of potential drugs for their effect on the expression of protein. However, the expertise and cost involved has so far limited its use as a routine technique. The overall workflow in 3D electrophoresis is summarized in Figure 10.

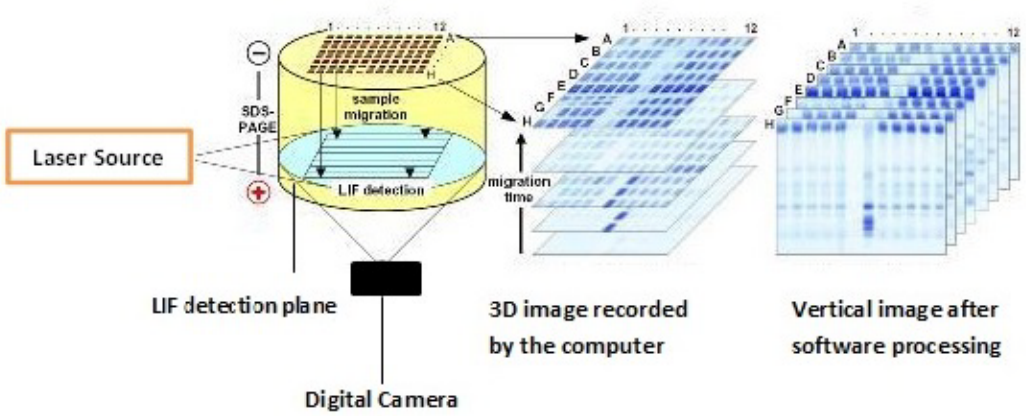


Figure 10: Workflow of 3D gel electrophoresis system. [Adapted from http://www.3d-gel.com].

Ultracentrifugation

Ultracentrifugation is referred to as centrifugation at very high speeds which may go as high as 2,000,000xg. Due to the enormous centrifugal force applied, it is possible to separate proteins and other macromolecules by this technique.

Density gradient centrifugation involves a linear concentration gradient of sugar (sucrose), glycerol, or commercially available silica based density gradient media like Percoll (a trademark owned by GE Healthcare) is generated in a tube such that the highest concentration is on the bottom and lowest on top. The concentration gradient also behaves as density gradient as the density also increases with concentration. The sample is layered on top and centrifuged at ultra-high speeds. The heavier molecules migrate towards the bottom faster than lighter ones. When the proteins move through a density gradient, they encounter liquid of increasing density and viscosity, which counteract the increasing centrifugal force and the proteins are segregated into sharp bands or zones. If the centrifugation is performed in the absence of sucrose, as particles move farther and farther from the center of rotation, without getting concentrated in sharp bands. Once the centrifugation is over, the gradient is then fractionated and different bands are collected.

The disadvantages in ultracentrifugation are cross-contaminations among fractions. This may sometimes result in relatively poor yields and low resolution. In addition, Takeuchi and Saheki [20] have reported the problem of removal of lipids and apoproteins during ultracentrifugation of lipoprotein particles, though it may not be considered as a disadvantage in general terms.

Acknowledgment

The author wishes to thank Dr. Fahim H. Khan and Dr. Haseeb Ahsan, for their support.

References

1. Albertsson PA (1985) History of aqueous polymer two-phase systems, in Partitioning in aqueous two-phase systems (Walter et al. eds.), Academic Press, Inc.
2. Lindqvist B, Storgards T (1955) Molecular-sieving Properties of Starch. Nature, Lond. 175: 511-512.
3. PORATH J, FLODIN P (1959) Gel filtration: a method for desalting and group separation. Nature 183: 1657-1659.
4. RAYMOND S, WEINTRAUB L (1959) Acrylamide gel as a supporting medium for zone electrophoresis. Science 130: 711.
5. Shapiro AL, Viñuela E, Maizel JV Jr (1967) Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochem Biophys Res Commun 28: 815-820.
6. Cuatrecasas P, Wilchek M, Anfinsen CB (1968) Selective enzyme purification by affinity chromatography. Proc Natl Acad Sci U S A 61: 636-643.
7. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250: 4007-4021.
8. Jorgenson JW, Lukacs KD (1981) Free-zone electrophoresis in glass capillaries. Clin Chem 27: 1551-1553.
9. Zwir-Ferenc A, Biziuk M (2006) Solid Phase Extraction Technique – Trends, Opportunities and Applications. Polish J. of Environ. Stud. 15: 677-690.
10. Porath J, Carlsson J, Olsson I, Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 258: 598-599.
11. Hjerten S (1973) Some general aspects of hydrophobic interaction chromatography. J. Chrom. A. 87: 325–331.
12. Camper DV, Viola RE (2009) Fully automated protein purification. Anal Biochem 393: 176-181.
13. Garfin DE (1990) One-dimensional gel electrophoresis. Methods Enzymol 182: 425-441.
14. Ros A, Faupel M, Mees H, Oostrum Jv, Ferrigno R, et al. (2002) Protein purification by Off-Gel electrophoresis. Proteomics 2: 151-156.
15. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.
16. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250: 4007-4021.
17. Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26: 231-243.
18. Ventzki R, Stegemann J (2003) High-throughput separation of DNA and proteins by three-dimensional geometry gel electrophoresis: feasibility studies. Electrophoresis 24: 4153-4160.
19. Ventzki R, Rüggeberg S, Leicht S, Franz T, Stegemann J (2007) Comparative 2-DE protein analysis in a 3-D geometry gel. Biotechniques 42: 271, 273, 275 passim.
20. Takeuchi N, Saheki S (1993) [Evaluation and problems of ultracentrifugal technique for separation and analysis of serum lipoproteins: comparison with other analytical methods]. Rinsho Byori 41: 750-758.

Advances in Protein Chemistry

Chapter: Protein Structure Databases and 3D Structure Prediction Tools

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published August, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Protein Structure Databases and 3D Structure Prediction Tools

Adeel Malik[1]*, Ahmad Firoz[2,3], Obaid Afzal[4], Vivekanand Jha[2,5] and Inho Choi[1]*

[1]School of Biotechnology, Yeungnam University, Republic of Korea
[2]Biomedical Informatics Center of ICMR, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India
[3]School of Chemistry and Biochemistry, Thapar University, Punjab, India
[4]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Jamia Hamdard (Hamdard University), New Delhi, India
[5]Department of Nephrology, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India

***Corresponding authors:** Adeel Malik, School of Biotechnology, 280 Daehak-ro, Gyeongsan, Gyeongbuk 712-749, Yeungnam University, Republic of Korea; E-mail: adeel@procarb.org

Inho Choi, School of Biotechnology, 280 Daehak-ro, Gyeongsan, Gyeongbuk 712-749, Yeungnam University, Republic of Korea; E-mail: inhochoi@ynu.ac.kr

1. Introduction

Large scale genome sequencing projects are generating tremendous quantity of protein sequences; however, complete understanding of the biological function of these proteins necessitates detailed knowledge about their structure and function [1]. Majority of the functions of an organism are mediated by proteins, and all these functions are generally, determined by three-dimensional (3D) structure of the proteins [2]. One of the common goals in biological sciences is to functionally characterize a protein sequence by solving an accurate 3D structure of the protein under study [3]. X-ray crystallography and NMR spectroscopy are the most powerful experimental methods to study the structure of a protein [4,5] and the improvement in these methods has led to an increase in the number of protein structures in PDB. Moreover, together with the enhancement of computational technologies in recent times and the progress of latest and powerful computer programs, it has become easier to predict new structure models based on the huge growth in the number of protein structures deposited in the PDB [4]. These protein structures have been extremely helpful in the refinement of experimental structures [6,7], the design of site-directed mutants [8], the characterization of molecular function and structure-based drug design [9].

Therefore, in view of the growing number of 3D structures and models, and their indispensable role in accurate functioning of the proteins, there is a need to develop various computational resources that can aid in the storage and analysis of these structures. PDB serves as a primary source in areas of structural biology along with other web-based protein structure databases which come in a large variety of types and levels of knowledge content. Some of them exhibiting general purpose interest cover all experimentally determined structures and provide valuable links, analyses, and graphical representations involving their 3D structure and biological function. Many other databases have attempted to organize 3D structures based on their folds as these can provide insights into their evolutionary relationships which might not be easy to identify from sequence comparison only. There are many servers that compare folds which are predominantly helpful for newly determined structures, and especially those having unknown function. The other more specialized databases deal with specific families, diseases and various structural features [10]. In addition to these databases for experimentally determined 3D structures, some databases aim at storing 3D models of proteins based on homology or comparative modeling.

Similarly, computational structure prediction methods offer important information for the large fraction of sequences whose structures are not experimentally determined so far. Among the various protein structure prediction methods, threading and comparative modeling depends on similarity across most of the modeled sequence and at least one known structure. However, in case of *de novo* or *Ab initio* methods, the structure is predicted from sequence alone and does not require any prior similarity at the fold level between the modeled sequence and any of the known structures. The aim of this chapter is to provide a comprehensive list of web based protein structure databases and state of the art 3D structure prediction tools. In the end, a brief overview of the CASP experiment is also provided.

2. 3D Structure Databases

Protein structure databases serve as a resource for a variety of experimentally determined protein structures. The main aim of the majority of these protein structure databases is to categorize and annotate the protein structures, thereby presenting

the biological society access to the experimental data in a constructive manner. Data incorporated in protein structure databases often consists of three dimensional coordinates as well as experimental details, such as unit cell dimensions and angles for x-ray crystallography resolved structures. Since PDB is the master protein 3D structure database, therefore, before highlighting some of the major protein structure databases, a brief historical background on PDB is provided in the next section.

2.1 Historical background of PDB

In 1968, a small but increasing number of protein structures determined by X-ray diffraction, and the newly existing molecular graphics display, known as the Brookhaven RAster Display (BRAD), to visualize these protein structures in 3D were the driving forces that led to the birth of the PDB. Later, with the support of Walter Hamilton a chemist at the Brookhaven National Laboratory, Edgar Meyer (Texas A&M University) began to write software to store atomic coordinate files in a universal format to make them accessible for geometric and graphical assessment. By 1971, one of Meyer's programs, SEARCH, facilitated researchers to distantly access information from the database to study protein structures offline [11]. SEARCH was helpful in enabling networking, thus marking the practical establishment of the PDB.

Soon after the death of Hamilton's in 1973, Tom Koetzle took over the direction of the PDB, and in 1974 the first PDB Newsletter was circulated to explain the details of data deposition and remote access. At this instant, only thirteen structures were prepared for distribution and four were pending. The PDB continued in Brookhaven until 1998, and in 1999 the PDB was transferred to the Research Collaboratory for Structural Bioinformatics (RCSB) [12], under the directorship of Helen M. Berman of Rutgers University [13].

In 2003, the Worldwide PDB (wwPDB) was formed which made PDB an international organization, with RCSB PDB, the Macromolecular Structure Database at the European Bioinformatics Institute (MSD-EBI) and PDB Japan (PDBj) at the Institute for Protein Research at Osaka University, Japan, acting as the founding members [14]. This laid the foundation for the PDB to remain a universal worldwide resource of structural biology data [15].

2.2 Protein structure databases

The structure databases are divided into two subsections based on whether it consists of experimentally determined structures or structure models. All these databases are listed in Table 1.

S. No.	Database	Link	Description
1.	PDB	http://www.rcsb.org/pdb/home/home.do	X-ray & NMR data for biological macromolecules
2.	PDBsum	http://www.ebi.ac.uk/pdbsum/	Pictorial representation of 3D structures
3.	CATH	http://www.cathdb.info/	Protein domain structures
4.	SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/	Familial and structural protein relationships
5.	MMDB	http://www.ncbi.nlm.nih.gov/structure	3D structures that are linked to NCBI Entrez
6.	ModBase	http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi	Comparative protein structure models
7.	SWISS-MODEL Repository	http://swissmodel.expasy.org/repository/?pid=smr01&zid=async	Annotated 3D comparative protein structure models
8.	Database of Macromolecular Movements	http://bioinfo.mbb.yale.edu/MolMovDB	Motions that occur in proteins and other macromolecules
9.	PROCARB	http://procarb.org/	3D structures of protein-carbohydrate complexes & comparative models of glycoproteins
10.	PMDB	http://mi.caspur.it/PMDB/	3D protein models obtained by structure prediction methods.
11.	PDBTM	http://pdbtm.enzim.hu/	Protein Data Bank of Transmembrane Proteins
12.	OPM	http://opm.phar.umich.edu/	OPM provides spatial arrangements of membrane proteins with respect to the hydrocarbon core of the lipid bilayer.

Table 1: List of various important protein structure and protein model databases.

2.2.1 Experimentally solved structure databases:

2.2.1.1 Protein Data Bank (PDB)

1. It is the chief source of structural data for biological macromolecules. PDB was founded at Brookhaven National Laboratories (BNL) [16] in 1971 as an archive for biological macromolecular crystal structures [17]. As of May 2013, there are 90206 biological macromolecular structures deposited in PDB [18]. The PDB database contains information regarding experimentally-determined structures of proteins, nucleic acids, and complex assemblies. The data obtained by X-ray crystallography or NMR spectroscopy is deposited globally by biologists and biochemists which is freely accessible to the world wide community. Each PDB file contains *xyz* coordinates of atoms and an entry in the PDB also includes information regarding the chemistry of the macromolecule, the small-molecule ligands, various particulars of the data collection and structure refinement, and some structural descriptors. Altogether, a characteristic PDB entry has about 400 *unique* items of data. The PDB file format was formulated in 1976 and is very simple, human readable as well as used by countless computer applications [19].

2.2.1.2 SCOP

The proteins in PDB have structural similarities with other proteins and, may share a common evolutionary source. Therefore, the Structural Classification of Proteins (SCOP) database [20] was created so that access to this information could be facilitated. Besides all the proteins in the current version of PDB, it also includes many proteins for which there are published descriptions but whose co-ordinates do not exist yet. The classification in SCOP is based on hierarchical levels

where the initial two levels, family and superfamily, illustrate close and distant evolutionary relationships where as the third fold describes geometrical relationships. The organization of proteins in SCOP has been created by visual examination and comparison of structures, which offers a possibility of most accurate and useful results keeping in mind the current limitations of purely automatic procedures. A protein domain represents the unit of classification. Small and medium sized proteins have a single domain and are, therefore, treated as a whole where as the domains in large proteins are generally classified separately [21].

For each entry, SCOP provides links to co-ordinates, images of the structure, interactive viewers, sequence data and literature references. The user can search the SCOP database by using two methods. The homology based search allows users to enter a sequence and get a list of structures to which it has significant levels of sequence similarity. The key word search method retrieves, matches from both the text of the SCOP database and the headers of PDB structure files [20]. The most update release (23 Feb 2009) contains 8221 PDB Entries, 110800 domains and 1 literature reference [http://scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-1.75].

2.2.1.3 CATH

The CATH (Class, Architecture, Topology, Homology) database [22] is a hierarchical classification of protein domain structures, using labor-intensive curation supported by a range of classification and prediction algorithms; for instance, structural comparison [23] and hidden-Markov model (HMM)-based methods [24]. Before splitting of constituent chains each protein structure is verified to make certain it meets the selection criteria. Consecutively, these chains are divided into one or more individual domains and then classified into homologous super families depending on their structure and function [25]. Top of the hierarchy is represented by the Class, or C-level, in which the domains are classified by their secondary structure content—i.e. Mostly alpha-helical (Class 1), mostly beta-sheet (Class 2), both alpha-helical and beta-sheet secondary structure elements (Class 3) or have very little secondary structure (Class 4).Inside each class, the domains are then classified based on their Architecture (A-level)—i.e. similarities in the arrangement of secondary structures in 3D space, which is further sub-divided into one or more topology, or fold groups (T-level), where the connectivity between these secondary structures is taken into account. Lastly, the domains are classified into their particular Homologous super families (H-level), based on the similarities in structure, sequence and/or function. Sequence clustering at the H-level creates sequence families at <35% sequence identity (S-level), <60% (O-level), <95% (L-level) and 100% (I level) [25]. CATH now includes 173536 domains, 1313 folds and 2626 super families.

2.2.1.4 PDBsum

PDBsum is a web-based database that aims to complement the data already available on protein and nucleic acid structures from various sources like CATH [22] SCOP [20] MMDB [26] and NDB [27] for nucleic acids, etc. The database provides a summary of the proteins, nucleic acids, ligands, water molecules and metals in each PDB file in addition to various analyses of their structural features [28]. These include the images of the structure, annotated plots of each protein chain's secondary structure, thorough structural analyses created by the PROMOTIF program, PROCHECK summary results and schematic representation of protein–ligand and protein–DNA interactions [29]. For each new structure deposited in the PDB, there is a corresponding summary in PDBsum which can be accessed by its four letter PDB code. The latest version of PDBsum contains 93,419 entries (Last update: 4 May, 2013).

2.2.1.5 PDBTM

PDBTM is a protein databank of transmembrane proteins with known structures, and aims to bring together all transmembrane proteins that are submitted in the PDB and to determine their membrane-spanning regions. The PDBTM database was created by scanning all PDB entries with the TMDET algorithm, which employs only structural information to find the most probable position of the lipid bilayer and to differentiate between transmembrane and globular proteins. The database is updated weekly as soon as the new PDB entries are available, by running the TMDET algorithm on every new PDB file. The PDBTM database can be considered as an expansion of the PDB database, since it contains added information for each PDB entry. The database is structured in the same way as the PDB; the entries are identified by their PDB code and are grouped in subdirectories according to the middle two characters of their codes. This database provides a useful resource for people interested in studying transmembrane proteins. For example, it could be used to assign whether a binding site is situated in the lipid or in the aqueous phase, which in turn is of significant value design a drug that binds to a certain part of a receptor [30]. The current version consists of 1891 transmembrane proteins, which include 1626 alpha and 258 beta proteins [http://pdbtm.enzim.hu/].

2.2.1.6 Database of Macromolecular Movements

Motions of proteins are involved in several biological functions such as catalysis, regulation of activity, transport of metabolites, formation of large assemblies and cellular locomotion. Many of the instances of protein structures solved in multiple conformations can now be studied in a database framework and are freely accessible [http://bioinfo.mbb.yale.edu/MolMovDB]. Protein motions are classified hierarchically initially on the basis of size (distinguishing between fragment, domain and subunit motions) and then based on the packing. Packing classification further divides motions into different categories (e.g. shear, hinge, other) which depends on whether or not they involve sliding over a constantly sustained and firmly packed interface. Additionally, the database gives some hint about the evidence behind each motion (i.e. the type of experimental information or whether the motion is inferred based on structural similarity) and efforts to illustrate numerous aspects of a motion in terms of a standardized nomenclature (e.g. the maximum rotation, the residue selection of a fixed core, etc.). As compared to an individual protein identifier, each entry in the database is indexed by a *unique motion*

identifier, since a single macromolecule can have a numerous motions and the same vital motion can be shared between diverse macromolecules. Also, each entry has links to graphics and movies describing the motion, frequently portraying a possible interpolated pathway [31].

2.2.1.7 Orientations of Proteins in Membranes (OPM)

OPM database is a compilation of transmembrane, monotopic and peripheral proteins derived from PDB whose spatial arrangements in the lipid bilayer have been theoretically calculated and compared with the experimental data. The OPM database allows various different types of analysis, for instance, sorting and searching of membrane proteins on the basis of their structural classification, species, destination membrane, numbers of transmembrane segments and subunits, numbers of secondary structures and the calculated hydrophobic thickness or tilt angle concerning the bilayer normal. OPM can be browsed either by searching proteins by their name or PDB ID or by sorting of proteins in tables for various specific categories like type, class, superfamily, family, destination membrane or biological source. An individual web page is created for each membrane protein complex, and coordinate files of all proteins with calculated membrane boundary planes are accessible for downloading individually for each protein or as a whole dataset. Currently, OPM has about 2172 proteins grouped into 393 superfamilies and 667 families representing 538 species [32].

2.2.1.8 PROCARB

PROCARB is an open-access database which consists of three separately functioning components, i.e., (i) Core PROCARB module, includes 3D structures of protein-carbohydrate complexes taken from PDB, (ii) Homology Models module, consisting of manually constructed 3D models of N-linked and O-linked glycoproteins by comparative modeling, and (iii) CBS-Pred prediction module, consists of web servers to predict carbohydrate-binding sites using single sequence or server-generated PSSM. In PROCARB, numerous pre-computed structural and functional properties of complexes are also included for rapid analysis. Particularly, information about function, secondary structure, solvent accessibility, hydrogen bonds and literature reference is included. Additionally, each protein in the database is mapped to Uniprot, Pfam, PDB, and so forth. Currently the PROCARB consists of 604 experimentally verified protein carbohydrate complexes and 26 N-linked and 20 O-linked models in which at least one experimentally verified glycosylation site was modeled [33]. Figure 1 shows a representative N-linked homology model of Human Lysosomal alpha-glucosidase with three experimentally verified glycosylation sites (Uniprot ID = P10253).

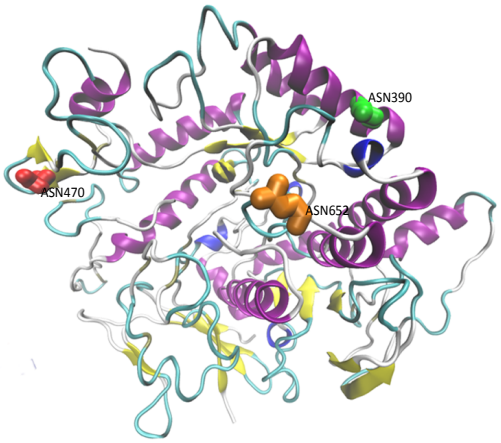


Figure 1: 3D model of Human Lysosomal alpha-glucosidase (Uniprot ID = P10253) with three experimentally verified N-linked glycosylation sites ASN390 (green), ASN470 (red) and ASN652 (orange).

2.2.2 Protein structure model databases

2.2.2.1 Protein Model Database (PMDB)

PMDB is a relational database of manually generated protein models deposited by users and attained with different structure prediction methods. The database provides easy access to models that have been published in the scientific literature, simultaneously with validating experimental data. Most of the models in PMDB are the predictions which have been submitted to the Critical Assessment of Techniques for Structure Prediction (CASP) experiment, as well as models generated by PMDB group, and the models uploaded based on published alignments [2]. For each protein target, one or more models could be available or several models for different regions of the same target protein. The database provides some information for each target and includes the protein name, sequence and length, organism and, whenever possible, links to the SwissProt sequence database. As soon as the structure of a target is determined, the PMDB database entry is also linked to the experimental structure in the PDB.

2.2.2.2 Molecular Modeling Database (MMDB)

MMDB offers simple access to the richness of 3D structure data and its huge potential for functional annotation [34]. MMDB reflects the contents of the PDB and is strongly integrated with NCBI's Entrez search and retrieval system. In MMDB, protein 3D structure data is connected with sequence data, sequence classification resources and PubChem, providing easy access to 3D structure data for structural biologists, as well as for molecular biologists and chemists [35]. An

entire set of comprehensive and pre-computed structural alignments are obtained with the VAST algorithm [36] where as the visualization tools for 3D structure and structure/sequence alignment are provided by the molecular graphics viewer Cn3D [37].

As on April 29, 2013, there are 89,571 structure records total which includes 22,084 proteins, 715 DNA and 508 RNA molecules only. Additionally, MMDB also consist of 2571 protein-DNA complexes, 1113 protein-RNA complexes and 116 protein-DNA-RNA complexes, in addition to more than 60,000 proteins bound to chemicals.

2.2.2.3 ModBase

ModBase is a database of annotated homology based protein structure models. Models in ModBase are generated as an automated software pipeline for comparative protein structure modeling, known as ModPipe [3] which mostly rely on modules of Modeller [38]. For fold assignment and target–template alignment, ModPipe uses sequence–sequence [39] sequence–profile [40,41] and profile–profile [40,42] methods by using an E-value cut-off of 1.0 to augment the possibility of identifying the finest available template structure. 10 models are generated [38] for each target–template alignment, and the model with the top Discrete Optimized Protein Energy (DOPE) statistical potential [43] score is selected and further assessed by numerous additional quality criteria: (i) target–template sequence identity, (ii) GA341 score [44] (iii) Z-DOPE score [43] (iv) ModPipe Quality Score (MPQS) and (v) TSVMod score [45].

Because of the rapid growth of the public sequence databases, models in ModBase are structured in data sets that are useful for specific projects. Currently, ModBase includes about 27,288,148 models and 4,332,658 *unique* sequences modeled for more than 50 complete genomes [46]. ModBase can be queried through its web interface by querying with UniprotKB [47] and GI [48] identifiers, gene names, annotation keywords, PDB codes [49] data set names, organism names, sequence similarity to the modeled sequences (BLAST [41]) and model-specific criteria such as model reliability, model size and target–template sequence identity. Additionally, the coordinate and alignment files can also be retrieved as text files [50].

2.2.2.4 SWISS-MODEL Repository

SWISS-MODEL Repository is a database of 3D protein structure models constructed by using the SWISS-MODEL homology-modelling pipeline based on protein sequences from the UniProt database [47]. The SWISS-MODEL pipeline integrates various steps like: template selection, target sequence and template structure alignment, model building, energy minimization and/or refinement and model quality assessment [51] Model target sequences are individually identified by their md5 cryptographic hash of the full length raw amino acid sequence which permits the redundancy in protein sequence databases entries to be reduced, and in turn assists cross-referencing with databases by means of different accession codes. The current SWISS-MODEL Repository release contains 3143784 model entries for 2286870 *unique* sequences in the UniProt database (2013_02).

The database could be queried for particular proteins by using diverse database accession codes (e.g. UniProt AC and ID, GenBank, IPI, Refseq) or directly by means of the protein amino acid sequence. For a particular query protein, a graphical outline demonstrating the segments for which models or experimental structures are available is shown. SWISS-MODEL Repository users can review the quality of the models in the database; search for alternative template structures, and construct models interactively by the use of SWISS-MODEL Workspace [52]. Repository is updated on a regular basis to reflect the growth of the sequence and structure databases.

3. 3D Structure Prediction

3.1 A brief history of molecular modelling

The first homology based model dates back to 1969 when a wire and plastic models of bonds and atoms of α-lactalbumin was constructed by using the coordinates of a hen's egg-white lysozyme and adjusting, physically, those amino acids that did not match the structure [53]. The two proteins exhibited 39% of sequence identity. Afterwards, the crystal structure of lysozyme was used to generate a model for α-lactalbumin [54]. These models were created by taking the existing coordinates of the well-known structure, and mutating side chains that were not identical in the protein to be modeled. This approach to protein modeling is still used at present with substantial success, particularly when the proteins share a considerable degree of sequence similarity [55].

McLachlan and Shotton [56] used the structures of mammalian chymotrypsin and elastase, and modeled the structure of α-lytic proteinase of the fungus Myxobacter 495. The modeling was not easy as the sequence similarity between the target and the template was only about 18%. Subsequently, the crystal structure of α-lytic proteinase was determined and compared with the homology model [57]. Although the domains of the model were constructed accurately, it was found that misalignment of the sequences led to local errors.

The modeling of variable regions was introduced in proteins on the basis of equivalent regions from homologous proteins of known structures [58,59]. Therefore, in order to construct the homology models of various serine proteases, structures of trypsin, chymotrypsin and elastase were superimposed, and it was found many equivalent Cα atoms were within 1.0Å of one another. The regions comprising of the amino acids of these Cα atoms were described as structurally conserved regions (SCRs). All the other remaining sites correspond to structurally variable or loop regions (VR) where the insertions/deletions were located. The backbone of SCRs and VRs was generated from the fragments of known serine proteases, where as the side chains were modeled based on the conformation found at the equivalent locations for those identical side chains in the well-known structures.

Furthermore, the initial models for the aspartic proteinases renin and renin-inhibitor complexes were built by using the 3D structure of the remotely related fungal proteinases [60-62]. Later on, the models for renin were constructed by employing the structures of mammalian aspartic proteases, pepsin and chymosin [63,64]. Comparative analysis of fungal and mammalian renin models revealed that the inaccuracies in the models occurred due to the dissimilarity in the arrangement of helices and strands between the fungal and mammalian proteinases, as well as the slightly different variable regions. On the other hand, the modeling of renin active site was reasonably accurate [65].

Early in the eighties, manual homology modeling was assisted by manoeuvring of protein molecules on the graphics terminal that was made achievable by computer programs like FRODO [66]. Since then, many different homology modeling packages have been developed [42], which can be grouped into three different groups: rigid-body assembly, segment matching, or modeling by satisfaction of spatial restraints [67].

3.2 Protein 3D structure prediction tools

The prediction of 3D structures of proteins remains an exceedingly complicated and uncertain undertaking. However, these difficulties can be addressed up to a certain extent by using some of the key state of the art tools which have been developed over the years. These tools (Table 2) either employ homology based methods or *Ab initio* methods in case of no significant similarities are found.

SNo.	Software	Link	Description
1.	MODELLER	http://salilab.org/modeller/	Satisfaction of spatial restraints
2.	SWISS-MODEL	http://swissmodel.expasy.org/	Local similarity/fragment assembly
3.	3D-JIGSAW	http://bmm.cancerresearchuk.org/~3djigsaw/	Fragment assembly
4.	ROBETTA	http://robetta.bakerlab.org/	Rosetta homology modeling and ab initio fragment assembly with Ginzu domain prediction
5.	RaptorX	http://raptorx.uchicago.edu/	Remote homology detection, protein 3D modeling, binding site prediction
6.	ESyPred3D	http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/	Template detection, alignment, 3D modeling
7.	HHpred	http://toolkit.tuebingen.mpg.de/hhpred	Template detection, alignment, 3D modeling
8.	EasyModeller	NA	GUI to MODELLER
9.	CPHModel	http://www.cbs.dtu.dk/services/CPHmodels/	Fragment assembly
10.	BHAGEERATH-H	http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp	Combination of *ab initio* folding and homology methods
11.	GeneSilico	https://genesilico.pl/meta2	'meta-server' for 3D structure prediction
12.	Geno3D	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html	Predicts 3D models based on distance geometry, simulated annealing and energy minimization algorithms.
13.	PSIPRED Protein Sequence Analysis Workbench	http://bioinf.cs.ucl.ac.uk/psipred/	Several high quality protein structure prediction and function annotation algorithms
14.	I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	Combination of *ab initio* folding and threading methods
15.	QUARK	http://zhanglab.ccmb.med.umich.edu/QUARK/	Monte Carlo fragment assembly
16.	MUSTER	http://zhanglab.ccmb.med.umich.edu/MUSTER/	Profile-profile alignment
17.	SPARKS-X	http://sparks.informatics.iupui.edu/yueyang/sparks-x/	3D structure modeling by Fold recognition according to Sequence profiles and structural profiles
18.	Phyre and Phyre2	http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index	Remote template detection, alignment, 3D modeling, multi-templates, *ab initio*
19.	PEP-FOLD	http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD/	*de novo* structure prediction of linear and disulfide bonded cyclic peptides

Table 2: List of protein 3D structure prediction tools.

3.2.1 Homology modeling

3D structure of a protein is capable of providing invaluable information about the function of a protein and allowing an efficient design of experiments, for instance site-directed mutagenesis, studies of disease-related mutations or the structure based drug designing efforts [68]. Traditional approaches to determine the 3D structure of a protein includes X-ray crystallography or NMR spectroscopy. Other theoretical methods have not shown much promise in providing high-resolution information for the bulk of proteins. The number of structurally characterized proteins is very less in comparison to the number of known protein sequences. As of May 07, 2013, there are 90,424 structures in PDB [http://www.rcsb.org/pdb/home/home.do] which is extremely low as compared to UniProtKB/Swiss-Prot (http://www.uniprot.org/) which contains 5,40,052 sequence entries as of May 01, 2013. Nevertheless, it seems quite unreasonable to believe that it is possible to experimentally determine the structures of all these hundreds and thousands of proteins regardless of immense growth in the efforts of structural genomics. Therefore, in view of the above, homology modeling (also known as comparative modeling) methods offer the only possible way to get structural information for such a huge number of proteins [69].

One of the prerequisites of successful model building requires the availability of at least one experimentally determined 3D structure known as template that shares a significant amino acid sequence similarity to the target sequence [68]. The main steps that are required to create a homology based model are summarized in Figure 2 and include: (1) identification of homologs that can be used as templates for modeling; (2) target-template sequence alignment; (3) building a model for the target based on the information from the alignments; and (4) evaluation of the model [70,71]. These modeling steps usually involve extensive expertise in structural biology and the use of extremely specialized computational tools [72]. Some of these highly specialized and frequently used homology based modeling tools are summarized below.

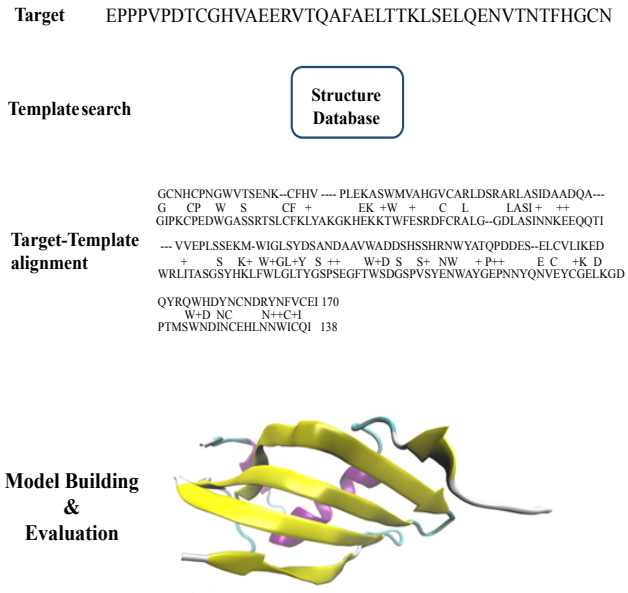


Figure 2: The target sequence is first searched against a structure database (preferably PDB) for an appropriate template. Target-template sequence alignment is created after the template with significant sequence similarity is found. The model is generated based on the target-template alignment and evaluated finally.

3.21.1 Modeller

Modeller is prediction software regularly used for creating homology or comparative protein structure models for proteins lacking experimentally determined 3D structures. It employs a method known as satisfaction of spatial restraints motivated by NMR spectroscopy data processing, by means of which a set of geometrical criteria are used to produce probability density functions (pdfs) for the location of each atom in the protein.

The Cα-Cα distances, main-chain N-O distances, main-chain and side-chain dihedral angles are restrained by these pdfs. In order to reduce the problem of a sparse database a smoothing method is used in the derivation of these relationships. The 3D model of a protein is acquired by optimization of the molecular pdf so that the model defies the input restraints as slightly as possible. The optimization method is a variable target function process that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. All these steps in Modeller are fully automated [38].

Modeller can perform additional tasks also, like *de novo* modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignments of protein sequences and/or structures, clustering, sequence database search and comparison of protein structures [73]. MODELLER was originally written in FORTRAN 90 and maintained at the University of California, San Francisco. It runs on UNIX, windows and MAC computers via scripts written in the Python language and is freely available for academic use. It does not provide any graphical interface however; a freely available GUI to MODELLER called Easy Modeller is developed and available for free download for Linux and Windows OS. A graphical user interfaces and commercial versions are also distributed as part of Accelrys' InsightII, and Discovery Studio interactive molecular modeling programs, which also contain many other tools for protein modeling and structural analysis. These programs facilitate preparation of input files for MODELLER as well as an analysis of the results.

3.21.2 SWISS-MODEL

SWISS-MODEL is an automated web server to build protein structure homology models, accessible via the ExPASy web server and program Deep View. The SWISS-MODEL provides a personalized web-based area for each user in which protein homology models can be built, and the results are stored and analyzed [74]. A personal account for the SWISS-MODEL workspace can be made and accessed at http://swissmodel.expasy.org/workspace. The user data is stored in a password-protected personal user space. For a given target protein in SWISS-MODEL, a library of experimental protein structures is searched to identify suitable templates. On the basis of a sequence alignment between the target protein and the template structure, a three-dimensional model for the target protein is generated. For template identification is carried out using various options like BLAST, PSI-BLAST and HMM-HMM-based searching. Automated mode can be applied if the alignment between the target and the template sequences exhibits a sufficiently high similarity where as alignment mode can be useful for distantly related target and template sequences. In case the sequence alignments fall within so-called 'twilight zone', i.e. when the sequence identity between target and template is lower than 30%, it is worthwhile to visually inspect and manually edit the target–template alignments. This leads to a considerable improvement in the quality of the resulting model and project mode can be applied [75]. To estimate the quality of model(s), programs provided in the 'Structure Assessment' section under tools using options sequence identity, stereochemistry check, global model quality and local model quality estimation are available. The quality assessment methods that are embedded in SWISS-MODEL workspace consists of Anolea mean force potential plots [76], GROMOS empirical force field energy [77], Verify3D profile

evaluation [78], Whatcheck [79] and Procheck [80], which produces reports so that the user can estimate the quality of protein models and template structures.

The SWISS-MODEL Repository is also available for the annotation of three-dimensional comparative protein structure models generated by homology-modeling pipeline SWISS-MODEL [51].

SWISS-MODEL is developed by the Computational Structural Biology Group at the Swiss Institute of Bioinformatics (SIB) and the Biozentrum of the University of Basel. SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and further developed at GWER - Glaxo Wellcome Experimental Research in Geneva and the SIB - Swiss Institute of Bioinformatics by Manuel Peitsch, Nicolas Guex and Torsten Schwede. Since 2001, SWISS-MODEL is being developed by Torsten Schwede's Structural Bioinformatics Group at the Biozentrum (University of Basel) and SIB - Swiss Institute of Bioinformatics.

3.21.3 3D-JIGSAW

3D-JIGSAW is a web-based server which generates 3D models for proteins based on homologues of known structure. The program can either be run in completely automatic mode by means of a web server or the individual modules of the program may be executed separately and the intermediate files saved which can be modified, if necessary, before the next module in the series is run [75]. In automatic mode, the program looks for homologous templates in its sequence databases and splits the query sequence into domains and the best covered domain is modeled using a maximum of two good templates, if found. The process may take up to an hour on a system. An e-mail with the alignment between query and template and a PDB formatted set of coordinates is sent to user, while in an interactive mode an e-mail is sent back to user with a link to a graphical display of the domain arrangement and useful information extracted from the PFAM database. From this link, user can select the domains needed to be modeled and can select the templates and correct the alignments before submitting a modeling job. Templates are ranked according to the coverage of the query, their sequence identity and their crystallographic resolution. Information from each template is easily accessed, including its alignment to the query sequence. In order to remove the steric clashes, the models are subjected to 100 steps of steepest descents energy minimization (unrestrained) by using the program CHARMM [81]. Additionally, error approximations were made by measuring the range of equivalent atomic displacements, as calculated from the superposition of all significant homologues [82]. A new version of 3D-JIGSAW (version 3.0) is also available where templates are identified using HMM [83] and the returned alignments are used to generate the models.

3.21.4 Robetta

Robetta web server offers automated structure prediction and analysis tools to deduce protein structural information from genomic data. The server utilizes the completely automated structure prediction process that constructs a model for a whole protein sequence whether the sequence homology to proteins of known structure is available or not. Robetta breaks down the input sequences into domains and generates models for domains with sequence homology to proteins of known structure using comparative modeling where as models for domains lacking such homology are constructed using the Rosetta *de novo* structure prediction method. Domain predictions and molecular coordinates of models across the full-length query are given as results [84].

The server can also exploit nuclear magnetic resonance (NMR) constraints data supplied by the user to resolve protein structures using the RosettaNMR [85,86] protocol. These tools can be used in combination with existing structural genomics initiatives to aid accelerate structure determination and expand structural insight for targeted open reading frames (ORFs). Robetta uses a somewhat modified edition of the *de novo* structure prediction method [87] so that the queries can be run within reasonable timescales. Similar to the original protocol, Robetta generates three- and nine-residue fragment libraries that correspond to local conformations seen in the PDB, and then assembles models by fragment insertion using a scoring function that favors protein-like characteristics. About 10000 decoys are generated for the original query where as 5000 decoys for up to two sequence homologs are generated by Robetta. Subsequently, 2000 query decoys and 1000 decoys of each homolog are selected on the basis of score, and on whether they surpass filters that remove decoys having too many local contacts or unlikely strand topologies. The selected decoys are then clustered based on Ca root-mean-square deviation (RMSD) above the entire ungapped positions. The top 9 cluster centers are selected as the top ranked models, and the finest scoring model that passed the filters is chosen as the 10th model. Searches with these final models are then carried out against a representative set of PDB chains to discover comparable structures using Mammoth [88] to recognize potential similarities to proteins of known structure.

In view of the fact that it is difficult to do *de novo* structure predictions with high accuracy, two moderately liberal definitions of correct predictions were defined. A domain is considered to be accurately modeled if at least 1 of the 10 models has a Mammoth alignment of 50 or more residues with an RMSD of 4Å or less to the native structure, or a Mammoth Z-score of 6 or better to the native structure. Robetta produces correct predictions for more than half of the domains on the basis of these definitions, especially with domains that consist of either all-alpha or alpha-beta secondary structure.

3.21.5 ESyPred3D

ESyPred3D is an automated homology modeling web server. It implements four steps of homology modeling approach which includes database searching to identify structure homolog, target template alignment, model building & optimization and model evaluation. ESyPred3D is available for use at http://www.fundp.ac.be/urbm/bioinfo/esypred/. The data and parameters required to submit the simplest job are the sequence, email id and description of job. In the advance parameter, PDB ID with chain name can be given to use this as template or PDB ID can be submitted to discard it as template. User has

the option to get results of submitted job in different format and alignment method provided at the web server. The method in EsyPred3D program gets benefit of the increased alignment performances of a new alignment strategy using neural networks, where alignments are obtained by combining, weighting and screening the results of several multiple alignment programs. The final three dimensional structures are built using the modeling package Modeller by the satisfaction of spatial and geometric restraints and an extremely rapid molecular dynamic annealing. Comparison has been made to test the accuracy of EsyPred3D models in CASP4 experiment and it was found that alignment strategy of EsyPred3D is better as compared to PSI-BLAST alignment and also models generated by EsyPred3D were among the best of CASP3 experiments [89].

3.21.6 CPHModel

In CPHModel, template recognition is done on the basis of profile-profile alignment guided by secondary structure and exposure predictions and in the new version, improvement in the alignment algorithm for the remote homology modeling has been introduced by structure dependant gap penalty [90]. The response time of the CPHModel web server is less than 20 minutes. Only one sequence per submission is allowed with not less than 15 and not more than 4,000 amino acids and the sequences are kept confidential and will be deleted after processing. In Query sequence section, the query sequence submitted and template hits obtained with the e-value are shown. In retrieving template section, if any significant hits were found in the PDB, the PDB entry name and the chain identifier are listed for the template that is used to construct the model. In making profile-profile alignment section, the resulting sequence alignments of using the PDB-Blast hit are shown. The results from the remote homology modeling is shown in another section and file with coordinates by clicking on the link 'query.pdb' can be downloaded in the in PDB format and finally if using a java-enabled browser, the trace of the model will be shown [90].

3.21.7 BHAGEERATH-H

BHAGEERATH-H is a homology-*Ab initio* hybrid web server for protein structure prediction, and is available at http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp. For each given input sequence, the web server predicts 5 structures. For sequences greater than 100 amino acids the results may take longer time to complete. In the template information section, user can also select auto template searching by using Bhageerath-H template search protocol. User can also give as many as references with option PDB ID and its chain ID by clicking ADD button. For sequences with homologs, it has the potential to predict a structure with higher resolution and accuracy in less time [91]. The models are ranked using all atom energy based empirical scoring function [92] and selects 100 lowest energy structures. The number of models are further reduced to five using solvent accessible surface areas (SASA) [http://www.bioinf.manchester.ac.uk/naccess/] which is further refined to five models by an optional short molecular dynamics simulations with explicit solvent.

3.21.8 EasyModeller

EasyModeller is a GUI standalone tool for homology modeling using MODELLER in the backend and was available initially for Windows at http://www.uohyd.ernet.in/modellergui/ and now Linux platform also. The GUI eliminates the requirement of prior knowledge of backend applications, consequently increasing the number of users of MODELLER and assists them to exploit the *unique* features of package more efficiently. EasyModeller uses default parameters for most commands to make the process as simple as possible. User can change the parameters manually by editing the associated script file in the working directory. It is freely available and its usage requires pre installed Modeller, perl, python, Microsoft Excel plot function and also Rasmol to view the modeled structure. User is required to follow the numbered steps one by one and is guided by the provided help panel. A very simple color coding consisting of green and red buttons is followed by EasyModeller where green represents the optional steps where as red are the essential steps to obtain a model. A maximum of six templates can be used in case of multi-template based modeling [93].

3.21.9 GeneSilico

GeneSilico is a WWW 'meta-server' which serves as a gateway to numerous protein structure prediction methods, and addresses the key issues of multiple sequence alignment (MSA) submission and data confidentiality [94]. The input to this meta-server could be either a single sequence or the alignment of the sequences. In case of single sequences each method generates its own MSA, whereas, if a MSA is submitted, the user can select between submissions of a full-length query or limit the analysis to a region with less than 30% gaps in the alignment. This allows elimination of highly divergent loops that may cause difficulty when the core structures of the template and the target are matched. The user-defined MSA is submitted to those fold recognition servers which allow the submission of MSA. For servers which allow the submissions of single sequences only and generate their own MSA, the user defined MSA is converted into a 'consensus sequence'. The consensus sequence can be generated by any of the alternate methods available. The quality of the target-template alignments obtained by GeneSilico meta-server ensured its win in the CASP5 homology modeling contest [95].

The various components that coordinately work in GeneSilico include: a) HMMPFAM: A tool for the identification of PFAM domains [96]; b) Secondary structure is predicted by using methods like PSIPRED [97], SAM-T02 [98] and PROF [99]; c) Identification of potential transmembrane helices is carried out by methods like MEMSAT2 [100], TMHMM [101], TMPRED [102]; d) A local PDB-BLAST filter is employed for the detection of closely related sequences of known structures in PDB; e) The FR methods comprising of RAPTOR [103], 3DPSSM [104], FUGUE [105], mGENTHREADER [106], FFAS [107], SAM-T02 [108] and BIOINBGU [109] represent the 3D structure prediction core; f) The results retrieved from these FR servers are analyzed using the consensus server PCONS [110], which ranks the best models generated by the FR methods; g) Based on the target-template alignments from FR methods, initial 3D models of the query structure are created using SCWRL [111], on the basis of the template backbone. These preliminary models do not contain the features

corresponding to gaps in the FR alignment (for example insertions in the target, absent from the template), however, the structure of the hydrophobic core is generally inferred sufficient to carry out 3D structure assessment using VERIFY3D [112]. Thus, all FR alignments achieved from diverse servers go through unified evaluation by energetic criteria employed in VERIFY3D in addition to the ranking criterion obtained by the PCONS server.

3.21.10 Geno3D

It is an automatic web server that uses distance geometry, simulated annealing and energy minimization algorithms to build the protein 3D models. The steps taken by the server to generate the homology models of the query sequence include: (a) identification of homologous proteins with known 3D structures by using PSI-BLAST; (b) in the second step all potential templates are provided to the user via a very handy user interface for target selection; (c) the alignment of both target and template sequences is performed; (d) the geometrical restraints (dihedral angles and distances) for corresponding atoms between the query and the template are extracted; (e) finally, the 3D models of the protein are constructed by using a distance geometry approach which are sent to the user by e-mail. The results which are sent to the user includes files containing atomic coordinates of each model (3 is the default value) that best satisfies the spatial restraints and the alignment between target and template sequences, and the superimposition (which highlights the poorly defined regions that correspond to gaps in the alignment) of the models with the template. In addition of these, a matrix of binaries root mean square deviations between coordinates of equivalent α-carbons is also sent to the user, in order to check the homogeneity of the models. The advantages in employing a distance geometry based approach in homology modeling are no *a priori* preference in loops construction and effortless recognition of well defined regions [113].

3.21.11 The PSIPRED Protein Sequence Analysis Workbench

The PSIPRED server maintained by the UCL Bioinformatics Group presents several high quality protein structure prediction and function annotation algorithms including PSIPRED- for the prediction of secondary structure (Jones, 1999), pGenTHREADER (fold recognition) and pDomTHREADER (domain recognition) [114], MEMSAT- transmembrane topology prediction [115,116], MetSite- metal binding sites [117], DISOPRED2- regions of protein disorder [118], DomPred- prediction of protein domain boundaries [119] and FFPred- protein function prediction [120] respectively. The web portal also offers a fully automated 3D modeling pipeline based on fragment-assembly approach, "BioSerf", which was placed in the top five servers in the *de novo* modeling category in CASP8 experiment [121].

3.2.2 Protein Threading

With the rapid growth in PDB and the improvement in structure prediction methods, Template-based modeling which includes homology modeling and protein threading is proving to be more powerful and key method for protein structure prediction. This suggests that the structures of several novel proteins can be predicted by means of template-based methods. The inaccuracy of a template-based model results from the selection of template and the alignment between sequence and template [122]. In case of higher sequence identity (>50%), template-based models can be accurately applied in virtual screening studies [123,124], designing site-directed mutagenesis experiments [125,126], small ligand docking prediction, and function prediction [1,127]. However, when sequence identity falls in the twilight zone (<30%), then it becomes difficult to identify the top most template and generate precise sequence-template alignments [128,129]. The alignments between the sequence and template can be significantly improved by integrating additional information in the form of sequence profile into the scoring function [122].

Methods such as HHpred [83], MUSTER [130], Phyre2 [131] and SPARKS/SP3/SP5 [132-136], are some of the prominent threading based methods which combine homologous information with a variety of structure based information for remote homolog detection.

3.2.2.1 HHpred

Functional information of a protein or gene can be predicted from homologous proteins or genes identified by various sequence based search methods like BLAST, FASTA or PSI-BLAST [41,137,138]. However, there are certain groups of proteins for which no orthologs have been indentified; therefore, it becomes quite difficult to establish a relationship to a protein with known function [83]. HHpred is one of the web-based servers which could be utilized when database search results from BLAST or PSI-BLAST retrieves insignificant hits against proteins having well known structures and functions [83]. For each query sequence in HHpred, an alignment of homologs is generated using PSI-BLAST search against non-redundant (NR) database of NCBI. PSIPRED [97] is used to predicted secondary structure and annotate the alignment from PSI-BLAST. This annotated multiple alignment is then used to generate a profile hidden Markov models (HMMs). Input to HHpred can be a single sequence or a multiple alignment. Search preferences include both local or global alignment and scoring secondary structure similarity. As an output HHpred produces pair wise query-template alignments, multiple alignments of the query with a set of templates selected from the search results, as well as 3D structural models generated by the MODELLER software from these alignments. HHpred also provides a high level of flexibility and user-friendliness while maintaining exceptional sensitivity.

HHpred is a web based server for remote protein homology detection and structure prediction and is the first to employ pair wise comparison of profile hidden Markov models (HMMs) and allows searching a variety of databases, like PDB, SCOP, Pfam, SMART, COGs and CDD. HHpred is very fast and performs uniformly well for single-domain as well as multi-domain query sequences and can be used to predict domain boundaries.

3.2.2.2 MUSTER

MUSTER (MUlti-Source ThreadER) is an extension of sequence profile–profile alignment (PPA) threading algorithm [139] which includes a variety of sequence and structural resources generated from several other tools. PPA algorithm has been effectively employed in tools like I-TASSER and LOMETS. MUSTER uses sequence and structure information like (1) sequence profiles; (2) secondary structures; (3) structure fragment profiles; (4) solvent accessibility; (5) dihedral torsion angles; (6) hydrophobic scoring matrix and merges them into single-body terms which can be easily used in dynamic programming search [130].

The major aim in this method is to thoroughly examine the increase that can be obtained in fold recognition when the different resources of structural features are combined with the powerful sequence profile–profile alignment methods. In order to identify the top match between the query and the template sequences, the Needleman-Wunsch [140] dynamic programming algorithm is used. The output of the MUSTER server includes top five template proteins and the query-template alignments and the models generated by MODELLER. The performance of MUSTER is improved over PPA because of the contribution of more accurate alignments by incorporating new structure features.

3.2.2.3 PHYRE

Phyre employs a library of well-known protein structures [131] retrieved from the Structural Classification of Proteins (SCOP) database [20] and increased with updated additions in the PDB. A profile is created and deposited in the fold library by scanning the sequence of all these structures against a non-redundant (NR) database. Additionally, the known and predicted secondary structure of these proteins is stored in the fold library as well.

A query sequence is scanned against the NR sequence database to construct profile in a similar manner. To accommodate together close and remote sequence homologs, five iterations of PSI-Blast are used. The resulting large numbers of pair wise alignments generated by PSI-Blast are pooled into a single alignment with the query sequence as the master. After the profile has been constructed, the three-state secondary structure (alpha helix (H), beta strand (E—for extended) and coil (C)) of the query is predicted using Psi-Pred [141] SSPro [142] and JNet [143].

Afterwards, the profile and predicted secondary structure is scanned against the fold library using a profile–profile alignment algorithm [144] which in turn returns a score on which the alignments are ranked. An E-value is generated by fitting these scores to an extreme value distribution. Finally, 3D models of the query are generated based on the top ten highest scoring alignments. Additionally, loop library and reconstruction procedure are used to repair missing or inserted regions caused by insertions and deletions in the alignment wherever possible. Moreover, by using a fast graph-based algorithm and side chain rotamer library, side chains are placed on the models. As an input, Phyre web server accepts an amino-acid query sequence and after about 30 minutes, a link related to the results including full downloadable 3D models of the query protein are sent to the user via email. The present publicly available Phyre server showed performance typical of the majority.

3.2.2.4 SPARKS-X

A chain of various single fold recognition methods like SPARKS, SP2, SP3, SP4 and SP5 were developed based on weighted matching of multiple profiles [145] including sequence profiles generated from multiple sequence alignment [41] predicted versus actual secondary structures [134,136,146] knowledge-based profile (single-body) score function [134] depth-dependent sequence profiles derived from template structures [136] predicted versus actual solvent accessible surface area [147] and predicted versus actual dihedral angles [133]. Among these, SPARKS, SP^3 and SP^4 were ranked amid the top performers for automatic servers in CASP 6 [135,148] and CASP 7 [147,149] experiments. However, a known concern in the above mentioned methods is that matching predicted 1D profiles of query sequence with actual profiles of templates is based on simple difference matrices.

These methods do not account for the likelihood of errors in predicted 1D structural properties; for instance secondary structure, backbone torsion angles and solvent accessible surface area. Therefore, in order to overcome these issues, energy terms based on the estimated probability of a match between predicted and actual 1D structural properties were introduced, a common technique used in fold recognition based on hidden Markov models [150]. This new web-based method is called as SPARKS-X and takes an additional advantage of newly improved accuracy in predicted secondary structure, torsion angles, mean absolute error and solvent accessibility [145]. The models are generated by modeler [38] based on the alignment created by SPARKS-X. In case of gaps of > 30 residues in the termini, the program will be evoked again to construct a model for the missing parts in the region. Subsequently, these different models are linked and steric clashes are removed by using the DFIRE potential functions [151,152].

SPARKS-X turned out to be one of the best single-method fold recognition servers as indicated by CASP 9 when tested for its alignment accuracy, fold recognition and structure prediction by using several benchmarks and compared it to several state-of-the-art techniques. The authors believe that the method can be further improved significantly by integrating the techniques of multiple templates and refinement in model building that are already being employed by many other automatic servers.

3.2.2.5 RAPTORX

RAPTOR – RApid Protein Threading by Operation Research technique is a novel linear programming approach to do predict 3D structure of proteins. In this approach, the protein threading dilemma is devised as a large scale integer

programming (IP) problem based on the contact map graph of the protein 3D structure template. Furthermore, the IP formulation is then relaxed to a linear programming (LP) problem, and later resolved by the canonical branch-and-bound scheme. RAPTOR extensively outperforms other programs at the fold similarity level as shown by benchmark tests for fold recognition. RAPTOR was ranked as top 1 among individual prediction servers, in terms of the recognition capability and alignment accuracy for Fold Recognition (FR) family targets as evaluated by CAFASP3 [153].

However, RAPTOR was significantly outperformed by RaptorX which is good at aligning proteins with sparse sequence profile. RaptorX utilizes a statistical knowledge based method to design a new threading scoring function, which aims at enhanced measuring the compatibility between a target sequence and a template structure. Additionally, RaptorX also has a multiple-template threading component and contains a new module for alignment quality prediction. RaptorX includes three main components: single-template threading, alignment quality prediction, and multiple-template threading. Initially, RaptorX aligns a specific target to each of its templates using the single-template alignment algorithm. Then, the alignment quality is predicted and all the templates are ranked by this predicted quality in a descending order. If the target is not appropriate for multiple-template threading, RaptorX generates a 3D model for the target from the pairwise alignment by means of the highest predicted quality. Else, RaptorX runs multiple-template threading for the target and constructs a corresponding 3D model. RaptorX does extremely well at the alignment of hard targets, which have less than 30% sequence identity with experimentally determined structures in PDB. RaptorX outperformed all the CASP9 participating servers including those using consensus and refinement methods, blindly tested on the 50 hardest CASP9 template-based modeling targets [154,155] As an input, RaptorX takes an amino acid sequence and predicts its secondary and tertiary structures as well as solvent accessibility and disordered regions. It also allocates confidence scores like P-value for the relative global quality, GDT (global distance test) and uGDT (un-normalized GDT) for the absolute global quality, and RMSD for the absolute local quality of each residue in the model to indicate the quality of a predicted 3D model [156].

3.2.3 Ab initio structure prediction

For a given protein amino acid sequence, the aim of *Ab initio* structure prediction is to predict the 3D structure of its native state. A protein sequence folds to a native conformation or a collection of conformations that is at or close to the global free-energy minimum. Therefore, the difficulty in finding native-like conformations for a given sequence has to address the development of a precise potential and an efficient method for searching the resultant energy landscape [157]. Traditionally, the most acclaimed structure prediction methods have been homology-based comparative modeling and fold recognition [158]. These two approaches construct protein models by aligning query sequences against solved template structures. If close templates are recognized, high-resolution models could be generated by the template-based modeling. However, if templates are lacking from the Protein Data Bank (PDB), the models require to be built from scratch, i.e. *Ab initio* folding, the most complicated category of protein-structure prediction [159,160]. As the size of the protein increases, the conformational phase space of sampling also increases sharply making the *Ab initio* modeling of bigger proteins exceedingly tricky [161]. Current *Ab initio* predictions are mainly focused on small proteins. Existing *Ab initio* predictions are primarily focused on small proteins and quite a few successful examples have been described in literature [139]. In this chapter, we highlight some of the well known current methods for predicting tertiary structure which employ *Ab initio* methods in the absence of homology to a known structure.

3.2.3.1 I-TASSER

I-TASSER is a web-based hierarchical protein structure modelling method which employs the Profile-Profile threading Alignment (PPA) improved by secondary-structure [162] and the iterative implementation of the Threading ASSEmbly Refinement (TASSER) program [163]. In I-TASSER, the query sequences are initially threaded through a representative PDB structure library (using 70% sequence identity as cut-off) to explore for the potential folds by four straightforward variants of PPA methods, with diverse combinations of the hidden Markov model [164] and PSI-BLAST [41] profiles and the Needleman-Wunsch [140] and Smith-Waterman [39] alignment algorithms. The continuous fragments are then eliminated from the threading aligned regions which are used to reconstruct full-length models whereas the threading unaligned regions (e.g. loops) are built by *Ab initio* modeling [165]. The replica-exchange Monte Carlo simulation [166] is used to search conformational space. SPICKER [167,168] is used to cluster the structure trajectories and the cluster centroids are acquired by the averaging the coordinates of the entire clustered structures.

Fragment assembly simulation is implemented again, which begins from the cluster centroid of the initial round simulation, so that the steric clashes on the centroid structures are excluded and promote additional model refinement. Spatial restraints are removed from the centroids and the PDB structures are explored by the structure alignment program TM-align [169], which are used to direct the second round of simulation. In the end, the structure decoys are clustered and the structures with the lowest possible energy in each cluster are selected, which has the Cα atoms and the side-chain centres' of mass specified. Backbone atoms (N, C, O) are added by Pulchra [170] and Scwrl_3.0 [171] is used to build side-chain rotamers. The absolute quality of the final model in comparison with the native structure is given by TM-score.

3.2.3.2 QUARK

QUARK is a web-based *Ab initio* protein structure prediction method which focuses on the elaborate design of the force field and the search engine by taking a semi-reduced model to represent protein residues by the full backbone atoms and the side-chain center of mass [172]. Initially, QUARK predicts a wide range of carefully selected structural features by using neural network (NN) for a query sequence. Based on the idea borrowed from Rosetta and I-TASSER, the global fold is then created by replica-exchange Monte Carlo (REMC) simulations by assembling the small fragments as generated

by gapless threading through template library. However, the fragments in QUARK have constantly multiple sizes which range from 1 to 20 residues where as in Rosetta and I-TASSER the fragments are in either 3/9-mer or from threading alignments. There are three main steps in which the QUARK *Ab initio* structure prediction method can be divided; A). Multiple feature predictions and fragment generation starting from one query sequence. B) Structural constructions using REMC simulation based on the semi-reduced protein model. C) Decoy structure clustering and full atomic refinement.

For each query amino acid sequence, multiple sequence alignment (MSA) is generated by PSI-BLAST [41] using a non-redundant database where as secondary structure (SS) is predicted by Protein Secondary Structure prediction program PSSpred [173] based on multiple neural network (NN) trainings of sequence profiles computed from the MSA. Features like Solvent accessibility (SA), real-value phi and psi angles, b-turn positions are predicted by using different NNs based on the checkpoint file by PSI-BLAST and SS types predicted by PSSpred. These features are then used to generate structural fragments for each segment of the query sequence. Moreover, the total energy of the QUARK force field is the sum of the 11 knowledge based energy terms that can be categorized into three levels of structural packing: atom-, residue-, and topology-level energy terms. The structural quality of the final models is measured by TM-score in comparison to the native structures. The QUARK method was also tested in the CASP9 experiment where it showed novel advancement in the field of *Ab initio* protein structure predictions.

3.2.3.3 PEP-FOLD

It is on-line resource for 3D structure prediction of peptides of length up to 36 amino acids with well-defined structures in aqueous solution and also accepts cyclic peptides using disulfide bonds defined by the user [174]. The PEP-FOLD predicts the structure in three steps. (i) Structural Alphabet (SA) letters from the amino acid sequence are predicted. The sequence is also used to generate a psi-blast profile which is further used as input for SVM that returns a likelihood profile of each SA letter at each position of the sequence. This SA profile is then analyzed to choose a few letters at each position. (ii) In the second step the 3D assembly of the prototype fragments linked with the letters selected is performed, which relies on the sOPEP coarse grained force field [175]. An improved greedy procedure [176] that builds the peptide residue by residue is utilized for 3D generation which is followed by a Monte-Carlo method for ultimate refinement. (iii) The final step creates all-atom conformations from the coarse grained models returned by the 100 simulations and executes a clustering procedure. At the end of the program run, the server gives the outcome of the cluster analysis. Additionally, if a reference structure was provided, the cRMSD, the GDT-TS and TM scores of each model are also reported. PEP-FOLD was tested on a benchmark of 34 cyclic peptides with one, two and three disulphide bonds and the best PEP-FOLD models deviated by an average RMS of 2.75Å from the full NMR structures. Similarly, a test on benchmark of 37 linear peptides showed that PEP-FOLD finds lowest-energy conformations deviating by 3Å RMS from the NMR rigid cores.

4. Critical Assessment of Protein Structure Prediction (CASP)

CASP is a community-wide, worldwide experiment for protein structure prediction which takes place after every two years ever since it started in 1994 [177]. CASP offers research groups with an opportunity to test their structure prediction algorithms and conveys an independent evaluation of the state of the art in protein structure modeling to the research community and software users. Currently, large number of servers and tools are extensively available for generating a structural model for the protein of interest which can then be used as a structural scaffold for designing additional experiments, interpreting functional data, assigning molecular function to the protein, or as a target for drug design or even as a means for solving the experimental structure of the protein. Nevertheless, the possible applications of a model depends on its quality, and therefore, the admittedly difficult problem of assessing in advance the quality of models created by various methods is of great significance [178].

CASP targets or the sequences of the protein for which experimental biologists are about to resolve a protein structure by X-ray crystallography or NMR spectroscopy are made available. Predictors generate and deposit models for these the CASP targets ahead of the structures are made publically available. If the given target sequence is found to exhibit a certain degree of similarity with known structure called template, then comparative protein modelling or homology based modeling may be used to predict the 3D structure or else, *de novo* protein structure prediction method must be applied, which is much less accurate but can occasionally yield models with the correct fold. A panel of three assessors compares the models with the structures the moment they are available and try to assess the quality of the models and to draw some conclusions about the state of the art of the diverse methods. To make certain that no predictor can have aforementioned information regarding a protein's structure, experiment is carried out in a double-blind fashion where neither the predictors nor the organizers and assessors know the structures of the target proteins at the time when predictions are made.

The main way of evaluation [179] is a comparison of the predicted model α-carbon positions with those in the target structure. The comparison is shown by cumulative plots of distances between pairs of corresponding α-carbon in the alignment of the model and the structure, and is assigned a numerical score GDT-TS (Global Distance Test — Total Score) [180] describing percentage of accurately-modeled residues in the model with regard to the target. Template-free, or *de novo modeling* (Free modeling) is also assessed visually by the assessors, given that the statistical scores do not work as well for finding loose similarities in the most complex cases [181].

In CASP10 [182] evaluation of the results was carried out in the following two main prediction categories:

4.1 Tertiary structure predictions (TS)

- *Template Based Modeling* (TBM) - includes domains where an appropriate template can be recognized that covers all

or almost the entire target.

- *Template free modeling* (FM) - includes models of proteins for which no reliable template can be identified.

- *Refinement* - one of the best models received selected during the prediction season was reissued as a starting structure for refinement.

- *Contact-assisted structure modeling* – this category demonstrates how the information of a few (usually 3 to 5) long-range contacts influences the capability of predictors to model the complete structure.

- *Chemical shifts guided modeling* of NMR structures was performed if the chemical shifts table from the NMR-spectroscopists was available for the selected CASP10 targets.

- *Structure modeling based on molecular replacement with Ab initio models and crystallographic diffraction data* was carried out for selected targets provided the structure factors from the crystallographers was available.

4.2 Other prediction categories

- Residue-residue contact prediction in proteins (RR).

- Disordered regions in target proteins (DR).

- Function prediction (prediction of binding sites) (FN).

- Quality assessment of models and the reliability of predicting certain residues in particular (QA).

The results of CASP are available in special supplement issues of the scientific journal *Proteins: Structure, Function* and *Bioinformatics*, all of which are accessible through the CASP website. CASP proceedings comprises of papers describing the structure and ways in which the experiments are conducted, the statistical evaluation procedures, reports from the evaluation teams highlighting state of the art in diverse prediction categories, methods from some of the most successful prediction teams, and advancement in different aspects of the modeling [182,183].

5. Molecular Visualization Tools

5.1 2D-GraLab (two-dimensional graphics lab for biosystem interactions)

2D-GraLab is a program that takes PDB file as an input and automatically generates schematic representations of nonbonding interactions across the protein binding interfaces. The program outputs two-dimensional PostScript diagrams providing intuitive and informative report of the protein–protein interactions (PPIs) and several energetics properties, such as hydrogen bond, salt bridge, van der Waals interaction, hydrophobic contact, π–π stacking, disulfide bond, desolvation effect, and loss of conformational entropy. The three main points on which 2D-GraLab emphasizes are a) reliability-which is ensured by the use of widely acclaimed programs embedded in 2D-GraLab; b) comprehensiveness- ability to handle nearly all the nonbonding interactions across binding interface of protein complexes, such as hydrogen bond, salt bridge, van der Waals (vdW) interaction, hydrophobic contact, π–π stacking, disulfide bond, desolvation effect, and loss of conformational entropy; c) artistry- with the aim of creating aesthetically pleasing 2D images of PPIs, the layout, color match, and page style for different diagrams are richly designed [183]. Additionally, 2D-GraLab also provides a graphical user interface (GUI), which allows users to interact with the program and displays the spatial structure and interfacial characteristics of protein complexes. 2D-GraLab is written in C++ and OpenGL, and the resultant output of 2D schematic diagrams of nonbinding interactions are depicted in PostScript. The current version of 2D-GraLab is freely available to academic users on request.

5.2 Chimera

UCSF Chimera is an extremely extensible program for interactive visualization and analysis of molecular structures, including density maps, supra-molecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles [184]. High-quality images and animations can be created by using Chimera which provides complete documentation and several tutorials, and is freely available for academic, government, non-profit, and personal use. Chimera is divided into two sections, namely, a *core* and *extensions*. The core offers fundamental services and molecular graphics capabilities, where as all the advanced level functionality is provided through extensions. The extension mechanism ensures that this design is strong enough to handle the requirements of outside researchers who wish to broaden the scope of Chimera in novel ways. These extensions can be incorporated into the Chimera menu system, and can present a separate graphical user interface as desired by means of the Tkinter (http://www.python.org/topics/tkinter/), Tix (http://tixlibrary.sourceforge.net/) and/or Pmw (http://pmw.sourceforge.net/) toolkits. The Chimera core consists of a C^{++} layer that handles time-critical operations such as graphics rendering, and a Python layer that handles all additional functions. The entire major C^{++} data and functions are made available to the Python layer. The abilities of core consist of molecular file input/output, molecular surface generation via the MSMS algorithm [185] and characteristics of graphical display for example wire-frame, ball-and-stick, ribbon, and sphere representations, transparency control, near and far clipping planes, and lenses.

Some of the major extensions of Chimera include:

Multiscale: This extension appends capabilities for interactively investigating bulky molecular assemblies with special focus on viral structures, condensed chromosomes, and ribosomes; further examples consist of cytoskeletal fibers and motors, flagellar structures, and chaperonins. Multiscale makes use of Chimera's core molecular display abilities, data

structures, file reading, and selection management, and the Volume Viewer extension for surface computation and rendering.

Multalign Viewer: This extension permits Chimera to show sequence alignments together with associated structures. Multalign Viewer can read and write sequence an alignment in a broad range of popular formats (Clustal, "aligned" FASTA, GCG MSF, GCG RSF, "aligned" NBRF/PIR, and Stockholm).

ViewDock: This extension assists in interactive screening of ligand orientations from DOCK [186], which calculates potential binding orientations given the structures of ligand and receptor molecules. ViewDock reads the DOCK output and gives a well-situated interface for screening results in the context of the target structure.

Collaboratory: Chimera's Collaboratory extension makes possible for researchers at geographically remote sites to share a molecular modeling session in real time. By default, all users associated to the identical session have equivalent control over the models (structures) being viewed. An alteration made by any member is straight away reflected to all other participants, so that a coordinated view of the data is sustained during the entire session.

Chimera is developed and maintained by the Resource for Biocomputing, Visualization, and Informatics and funded by the NIH National Center for Research Resources, and one of their major challenges is to enhance Chimera's performance in the context of the large-scale systems for which the Multiscale extension was created.

5.3 VMD: Visual molecular dynamics

VMD is a molecular graphics program intended for the display and analysis of molecular assemblies like proteins and nucleic acids. One of the most important features of VMD is that it can simultaneously display several numbers of structures by means of a broad array of rendering styles and coloring methods. Molecules are shown as one or more "representations," in which every representation symbolizes a particular rendering technique and coloring scheme for a selected subset of atoms [187]. The atoms displayed in each representation are selected by using a wide atom selection syntax, which includes Boolean operators and regular expressions. Besides providing a complete graphical user interface for program control, VMD also has a text interface which uses the Tcl embeddable parser to permit complex scripting tasks with variable substitution, control loops, and function calls. VMD has also been explicitly designed with the facility to animate molecular dynamics (MD) simulation trajectories, which could either be imported from files or from a direct connection to a running MD simulation. VMD is the visualization module of MDScope [188], a set of tools for interactive problem solving in structural biology, which also comprises the parallel MD program NAMD [189], and the MDCOMM software used to unite the visualization and simulation programs. VMD is written in C^{++}, using an object-oriented design; the program, including source code and all the documentation, is freely obtainable via anonymous ftp and through the World Wide Web.

5.4 Pymol

Pymol, a molecular visualization system developed by Warren Lyford DeLano and commercialized by DeLano Scientific LLC (a private software company committed to develop useful tools for scientific and educational communities), is a widely popular open source molecular visualization tool. Pymol can create superior quality 3D images of small molecules and biological macromolecules, such as proteins and nucleic acids. The Py segment of the software's name refers to the fact that it extends, and is extensible by the Python programming language. More details about Pymol can be found at http://www.pymol.org/.

Appendix

Section	Database/Software	Source
2.2.1.1	PDB	http://www.rcsb.org/pdb/home/home.do
2.2.1.2	SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
2.2.1.3	CATH	http://www.cathdb.info/
2.2.1.4	PDBsum	http://www.ebi.ac.uk/pdbsum/
2.2.1.5	PDBTM	http://pdbtm.enzim.hu/
2.2.1.6	Database of Macromolecular Movements	http://www.molmovdb.org/
2.2.1.7	Orientations of Proteins in Membranes (OPM)	http://opm.phar.umich.edu/
2.2.1.8	PROCARB	http://www.procarb.org/
2.2.2.1	Protein Model Database (PMDB)	http://mi.caspur.it/PMDB/
2.2.2.2	Molecular Modeling Database (MMDB)	http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml
2.2.2.3	ModBase	http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi
2.2.2.4	SWISS-MODEL Repository	http://swissmodel.expasy.org/repository/
3.2.1.1	MODELLER	http://salilab.org/modeller/
3.2.1.2	SWISS-MODEL	http://swissmodel.expasy.org/
3.2.1.3	3D-JIGSAW	http://bmm.cancerresearchuk.org/~3djigsaw/
3.2.1.4	Robetta	http://robetta.bakerlab.org/
3.2.1.5	ESyPred3D	http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred
3.2.1.6	CPHModel	http://www.cbs.dtu.dk/services/CPHmodels/
3.2.1.7	BHAGEERATH-H	http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp
3.2.1.8	EasyModeller	http://www.uohyd.ernet.in/modellergui/
3.2.1.9	GeneSilico	https://genesilico.pl/meta2/

3.2.1.10	Geno3D	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html
3.2.1.11	The PSIPRED Protein Sequence Analysis Workbench	http://bioinf.cs.ucl.ac.uk/psipred/
3.2.2.1	HHpred	http://toolkit.tuebingen.mpg.de/hhpred
3.2.2.2	MUSTER	http://zhanglab.ccmb.med.umich.edu/MUSTER/
3.2.2.3	PHYRE	http://www.sbg.bio.ic.ac.uk/~phyre/
3.2.2.4	SPARKS-X	http://sparks.informatics.iupui.edu/sparks-x/
3.2.2.5	RAPTORX	http://raptorx.uchicago.edu/
3.2.3.1	I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
3.2.3.2	QUARK	http://zhanglab.ccmb.med.umich.edu/QUARK/
3.2.3.3	PEP-FOLD	http://bioserv.rpbs.univ-paris-diderot.fr/PEP-FOLD/
4	CRITICAL ASSESSMENT of PROTEIN STRUCTURE PREDICTION (CASP)	http://www.predictioncenter.org/casp10/
5.1	2D-GraLab	http://2d-gralab.soft112.com/
5.2	CHIMERA	http://www.cgl.ucsf.edu/chimera/
5.3	VMD: Visual molecular dynamics	http://www.ks.uiuc.edu/Research/vmd/
5.4	PYMOL	http://www.pymol.org/

References

1. Baker D, Sali A (2001) Protein structure prediction and structural genomics. Science 294: 93-96.
2. Castrignanò T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A (2006) The PMDB Protein Model Database. Nucleic Acids Res 34: D306-309.
3. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, et al. (2003) Tools for comparative protein structure modeling and analysis. Nucleic Acids Res 31: 3375-3380.
4. Ilari A, Savino C (2008) Protein structure determination by x-ray crystallography. Methods Mol Biol 452: 63-87.
5. Göbl C, Tjandra N (2012) Application of Solution NMR Spectroscopy to Study Protein Dynamics. Entropy 14: 581-598.
6. Giorgetti A, Raimondo D, Miele AE, Tramontano A (2005) Evaluating the usefulness of protein structure models for molecular replacement. Bioinformatics 21 Suppl 2: ii72-76.
7. Jones DT (2001) Evaluating the potential of using fold-recognition models for molecular replacement. Acta Crystallogr D Biol Crystallogr 57: 1428-1434.
8. Laskowski RA, Watson JD, Thornton JM (2003) From protein structure to biochemical function? J Struct Funct Genomics 4: 167-177.
9. Jacobson M, Sali A (2004) Comparative protein Structure Modelling and its applications to drug discovery. Annual Reports in Medicinal Chemistry 39: 259–274.
10. Laskowski RA (2010) Protein structure databases. Methods Mol Biol 609: 59-82.
11. Meyer EF (1997) The first years of the Protein Data Bank. Protein Sci 6: 1591-1597.
12. RCSB: [http://home.rcsb.org/]
13. RCSB PDB Newsletter Archive: [http://www.rcsb.org/pdb/static.do?p=general_information/news_publications/newsletters/newsletter.html]
14. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. Nat Struct Biol 10: 980.
15. Berman HM (2008) The Protein Data Bank: a historical perspective. Acta Crystallogr A 64: 88-95.
16. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, et al. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112: 535-542.
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235-242.
18. http://www.rcsb.org/pdb/home/home.do
19. Berman HM (2008) The Protein Data Bank: a historical perspective. Acta Crystallogr A 64: 88-95.
20. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247: 536-540.
21. Hubbard TJ, Ailey B, Brenner SE, Murzin AG, Chothia C (1999) SCOP: a Structural Classification of Proteins database. Nucleic Acids Res 27: 254-256.
22. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, et al. (1997) CATH--a hierarchic classification of protein domain structures. Structure 5: 1093-1108.
23. Redfern OC, Harrison A, Dallman T, Pearl FM, Orengo CA (2007) CATHEDRAL: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. PLoS Comput Biol 3: e232.
24. Sillitoe I, Dibley M, Bray J, Addou S, Orengo C (2005) Assessing strategies for improved superfamily recognition. Protein Sci 14: 1800-1810.
25. Sillitoe I, Cuff AL, Dessailly BH, Dawson NL, Furnham N, et al. (2013) New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res 41: D490-498.
26. Chen J, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, et al. (2003) MMDB: Entrez's 3D-structure database. Nucleic Acids Res 31: 474-477.
27. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, et al. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J 63: 751-759.
28. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, et al. (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci 22: 488-490.
29. Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Res 29: 221-222.
30. Tusnády GE, Dosztányi Z, Simon I (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 33: D275-278.
31. Gerstein M, Krebs W (1998) A database of macromolecular motions. Nucleic Acids Res 26: 4280-4290.
32. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI (2006) OPM: orientations of proteins in membranes database. Bioinformatics 22: 623-625.
33. Malik A, Firoz A, Jha V, Ahmad S (2010) PROCARB: A Database of Known and Modelled Carbohydrate-Binding Protein Structures with Sequence-Based Prediction Tools. Adv Bioinformatics.

34. Wang Y, Addess KJ, Chen J, Geer LY, He J, et al. (2007) MMDB: annotating protein sequences with Entrez's 3D-structure database. Nucleic Acids Res 35: D298-300.

35. Madej T, Addess KJ, Fong JH, Geer LY, Geer RC, et al. (2012) MMDB: 3D structures and macromolecular interactions. Nucleic Acids Res 40: D461-464.

36. Gibrat JF, Madej T, Bryant SH (1996) Surprising similarities in structure comparison. Curr Opin Struct Biol 6: 377-385.

37. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH (2000) Cn3D: sequence and structure views for Entrez. Trends Biochem Sci 25: 300-302.

38. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234: 779-815.

39. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147: 195-197.

40. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2007) Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci Chapter 2: Unit 2.

41. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

42. Marti-Renom MA, Madhusudhan MS, Sali A (2004) Alignment of protein sequences by their profiles. Protein Sci 13: 1071-1087.

43. Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15: 2507-2524.

44. Melo F, Sali A (2007) Fold assessment for comparative protein structure modeling. Protein Sci 16: 2412-2426.

45. Eramian D, Eswar N, Shen MY, Sali A (2008) How well can the accuracy of comparative protein structure models be predicted? Protein Sci 17: 1881-1893.

46. http://salilab.org/modbase/statistics

47. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, et al. (2005) The Universal Protein Resource (UniProt). Nucleic Acids Res 33: D154-159.

48. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2010) GenBank. Nucleic Acids Res 38: D46-51.

49. Henrick K, Feng Z, Bluhm WF, Dimitropoulos D, Doreleijers JF, et al. (2008) Remediation of the protein data bank archive. Nucleic Acids Res 36: D426-433.

50. Pieper U, Webb BM, Barkan DT, Schneidman-Duhovny D, Schlessinger A, et al. (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 39: D465-474.

51. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37: D387-392.

52. http://swissmodel.expasy.org/workspace/

53. Browne WJ, North AC, Phillips DC, Brew K, Vanaman TC, et al. (1969) A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. J Mol Biol 42: 65-86.

54. Warme PK, Momany FA, Rumball SV, Tuttle RW, Scheraga HA (1974) Computation of structures of homologous proteins. Alpha-lactalbumin from lysozyme. Biochemistry 13: 768-782.

55. May AC, Blundell TL (1994) Automated comparative modelling of protein structures. Curr Opin Biotechnol 5: 355-360.

56. McLachlan AD, Shotton DM (1971) Structural similarities between alpha-lytic protease of Myxobacter 495 and elastase. Nat New Biol 229: 202-205.

57. Brayer GD, Delbaere LT, James MN (1979) Molecular structure of the alpha-lytic protease from Myxobacter 495 at 2.8 Angstroms resolution. J Mol Biol 131: 743-775.

58. Greer J (1980) Model for haptoglobin heavy chain based upon structural homology. Proc Natl Acad Sci U S A 77: 3393-3397.

59. Greer J (1981) Comparative model-building of the mammalian serine proteases. J Mol Biol 153: 1027-1042.

60. Carlson W, Handschumacher M, Karplus M, Haber E (1984) Studies of the three-dimensional structure of human renin and its inhibitors. J Hypertens Suppl 2: S281-284.

61. Carlson W, Karplus M, Haber E (1985) Construction of a model for the three-dimensional structure of human renal renin. Hypertension 7: 13-26.

62. Akahane K, Umeyama H, Nakagawa S, Moriguchi I, Hirose S, et al. (1985) Three-dimensional structure of human renin. Hypertension 7: 3-12.

63. Frazao C, Topham C, Dhanaraj V, Blundell TL (1994) Comparative modelling of human renin: A retrospective evaluation of the model with respect to the X-ray crystal structure. Pure & Appl. Chem. 66: 43-50.

64. Hutchins C, Greer J (1991) Comparative modeling of proteins in the design of novel renin inhibitors. Crit Rev Biochem Mol Biol 26: 77-127.

65. Azim MK 1999: http://eprints.hec.gov.pk/937/1/671.html.htm

66. Jones TA (1978) A graphics model building and refinement system for macromolecules. J. Appl. Cryst. 11: 268-72.

67. Wallner B, Elofsson A (2005) All are not equal: a benchmark of different homology modeling programs. Protein Sci 14: 1315-1327.

68. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 31: 3381-3385.

69. Wallner B, Elofsson A (2005) All are not equal: a benchmark of different homology modeling programs. Protein Sci 14: 1315-1327.

70. Sánchez R, Sali A (1997) Advances in comparative protein-structure modelling. Curr Opin Struct Biol 7: 206-214.

71. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, et al. (2000) Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 29: 291-325.

72. Tramontano A, Leplae R, Morea V (2001) Analysis and assessment of comparative modeling predictions in CASP4. Proteins Suppl 5: 22-38.

73. http://salilab.org/modeller/manual/%20

74. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.

75. Bates PA, Kelley LA, MacCallum RM, Sternberg MJ (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins Suppl 5: 39-46.

76. Melo F, Feytmans E (1998) Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 277: 1141-1152.

77. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH (1996) Biomolecular Simulations: The GROMOS96 Manual and User Guide. Zürich: VdF Hochschulverlag ETHZ.

78. Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277: 396-404.

79. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381: 272.

80. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK-a program to check the stereochemical quality of protein structures. J. Appl. Cryst 26: 283–291.

81. Wesson L, Eisenberg D (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1: 227-235.

82. Bates PA, Sternberg MJ (1999) Model building by comparison at CASP3: using expert knowledge and computer automation. Proteins Suppl 3: 47-54.

83. Söding J (2005) Protein homology detection by HMM-HMM comparison. Bioinformatics 21: 951-960.

84. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32: W526-531.

85. Bowers PM, Strauss CE, Baker D (2000) De novo protein structure determination using sparse NMR data. J Biomol NMR 18: 311-318.

86. Meiler J, Baker D (2003) Rapid protein fold determination using unassigned NMR data. Proc Natl Acad Sci U S A 100: 15404-15409.

87. Chivian D, Kim DE, Malmström L, Bradley P, Robertson T, et al. (2003) Automated prediction of CASP-5 structures using the Robetta server. Proteins 53 Suppl 6: 524-533.

88. Ortiz AR, Strauss CE, Olmea O (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. Protein Sci 11: 2606-2621.

89. Lambert C, Léonard N, De Bolle X, Depiereux E (2002) ESyPred3D: Prediction of proteins 3D structures. Bioinformatics 18: 1250-1256.

90. Nielsen M, Lundegaard C, Lund O, Petersen TN (2010) CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. Nucleic Acids Res 38: W576-581.

91. Jayaram B, Dhingra P, Lakhani B, Shekhar S (2012) Bhageerath-Targeting the near impossible: Pushing the frontiers of atomic models for protein tertiary structure prediction. J. Chem. Sci 124: 83–91.

92. Narang P, Bhushan K, Bose S, Jayaram B (2006) Protein structure evaluation using an all-atom energy based empirical scoring function. J Biomol Struct Dyn 23: 385-406.

93. Kuntal BK, Aparoy P, Reddanna P (2010) EasyModeller: A graphical interface to MODELLER. BMC Res Notes 3: 226.

94. Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res 31: 3305-3307.

95. Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, et al. (2003) A "FRankenstein's monster" approach to comparative modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. Proteins 53: 369-379.

96. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Barrell D, et al. (2003) The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res 31: 315-318.

97. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292: 195-202.

98. Karplus K, Karchin R, Barrett C, Tu S, Cline M, et al. (2001) What is the value added by human intervention in protein structure prediction? Proteins Suppl 5: 86-91.

99. Ouali M, King RD (2000) Cascaded multiple classifiers for secondary structure prediction. Protein Sci 9: 1162-1176.

100. Jones DT, Taylor WR, Thornton JM (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry 33: 3038-3049.

101. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175-182.

102. Hofmann K and Stoffel W (1993) TMbase—a database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler 374, 166.

103. Xu J, Li M, Kim D, Xu Y (2003) RAPTOR: optimal protein threading by linear programming. J Bioinform Comput Biol 1: 95-117.

104. Kelley LA, MacCallum RM, Sternberg MJ (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 299: 499-520.

105. Shi J, Blundell TL, Mizuguchi K (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 310: 243-257.

106. Jones DT (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 287: 797-815.

107. Rychlewski L, Jaroszewski L, Li W, Godzik A (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci. 9: 232–241.

108. Karplus K, Karchin R, Barrett C, Tu S, Cline M, et al. (2001) What is the value added by human intervention in protein structure prediction? Proteins Suppl 5: 86-91.

109. Fischer D (2000) Hybrid fold recognition: combining sequence derived properties with evolutionary information. Pac Symp Biocomput .

110. Lundström J, Rychlewski L, Bujnicki J, Elofsson A (2001) Pcons: a neural-network-based consensus predictor that improves fold recognition. Protein Sci 10: 2354-2362.

111. Dunbrack RL Jr (1999) Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL. Proteins Suppl 3: 81-87.

112. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356: 83-85.

113. Combet C, Jambon M, Deléage G, Geourjon C (2002) Geno3D: automatic comparative molecular modelling of protein. Bioinformatics 18: 213-214.

114. Lobley A, Sadowski MI, Jones DT (2009) pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics 25: 1761-1767.

115. Jones DT (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics 23: 538-544.

116. Nugent T, Jones DT (2009) Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics 10: 159.

117. Sodhi JS, Bryson K, McGuffin LJ, Ward JJ, Wernisch L, et al. (2004) Predicting metal-binding site residues in low-resolution structural models. J Mol Biol 342: 307-320.

118. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT (2004) The DISOPRED server for the prediction of protein disorder. Bioinformatics 20: 2138-2139.

119. Bryson K, Cozzetto D, Jones DT (2007) Computer-assisted protein domain boundary prediction using the DomPred server. Curr Protein Pept Sci 8: 181-188.

120. Lobley AE, Nugent T, Orengo CA, Jones DT (2008) FFPred: an integrated feature-based function prediction server for vertebrate proteomes. Nucleic Acids Res 36: W297-302.

121. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, et al. (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res 38: W563-568.

122. Peng J, Xu J (2010) Low-homology protein threading. Bioinformatics 26: i294-300.

123. Bjelic S, Aqvist J (2004) Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. Biochemistry 43: 14521-14528.

124. Caffrey CR, Placha L, Barinka C, Hradilek M, Dostál J, et al. (2005) Homology modeling and SAR analysis of Schistosoma japonicum cathepsin D (SjCD) with statin inhibitors identify a unique active site steric barrier with potential for the design of specific inhibitors. Biol Chem 386: 339-349.

125. Skowronek KJ, Kosinski J, Bujnicki JM (2006) Theoretical model of restriction endonuclease HpaI in complex with DNA, predicted by fold recognition and validated by site-directed mutagenesis. Proteins 63: 1059-1068.

126. Wells GA, Birkholtz LM, Joubert F, Walter RD, Louw AI (2006) Novel properties of malarial S-adenosylmethionine decarboxylase as revealed by structural modelling. J Mol Graph Model 24: 307-318.

127. Skolnick J, Fetrow JS, Kolinski A (2000) Structural genomics and its importance for gene function analysis. Nat Biotechnol 18: 283-287.

128. Chakravarty S, Godbole S, Zhang B, Berger S, Sanchez R (2008) Systematic analysis of the effect of multiple templates on the accuracy of comparative models of protein structure. BMC Struct Biol 8: 31.

129. Sánchez R, Pieper U, Melo F, Eswar N, Martí-Renom MA, et al. (2000) Protein structure modeling for structural genomics. Nat Struct Biol 7 Suppl: 986-990.

130. Wu S, Zhang Y (2008) MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. Proteins 72: 547-556.

131. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4: 363-371.

132. Zhang C, Liu S, Zhou H, Zhou Y (2004) An accurate, residue-level, pair potential of mean force for folding and binding based on the distance-scaled, ideal-gas reference state. Protein Sci 13: 400-411.

133. Zhang W, Liu S, Zhou Y (2008) SP5: improving protein fold recognition by using torsion angle profiles and profile-based gap penalty model. PLoS One 3: e2325.

134. Zhou H, Zhou Y (2004) Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. Proteins 55: 1005-1013.

135. Zhou H, Zhou Y (2005) SPARKS 2 and SP3 servers in CASP6. Proteins 61 Suppl 7: 152-156.

136. Zhou H, Zhou Y (2005) Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins 58: 321-328.

137. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

138. Pearson WR (1991) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics 11: 635-650.

139. Wu S, Skolnick J, Zhang Y (2007) Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol 5: 17.

140. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48: 443-453.

141. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. Bioinformatics 16: 404-405.

142. Pollastri G, Przybylski D, Rost B, Baldi P (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. Proteins 47: 228-235.

143. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36: W197-201.

144. Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins 70: 611-625.

145. Yang Y, Faraggi E, Zhao H, Zhou Y (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. Bioinformatics 27: 2076-2082.

146. Rost B, Schneider R, Sander C (1997) Protein fold recognition by prediction-based threading. J Mol Biol 270: 471-480.

147. Liu S, Zhang C, Liang S, Zhou Y (2007) Fold recognition by concurrent use of solvent accessibility and residue depth. Proteins 68: 636-645.

148. Tress M, Ezkurdia I, Graña O, López G, Valencia A (2005) Assessment of predictions submitted for the CASP6 comparative modeling category. Proteins 61 Suppl 7: 27-45.

149. Battey JN, Kopp J, Bordoli L, Read RJ, Clarke ND, et al. (2007) Automated server predictions in CASP7. Proteins 69 Suppl 8: 68-82.

150. Hargbo J, Elofsson A (1999) Hidden Markov models that use predicted secondary structures for fold recognition. Proteins 36: 68-76.

151. Yang Y, Zhou Y (2008) Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. Protein Sci 17: 1212-1219.

152. Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 11: 2714-2726.

153. Xu J, Li M, Lin G, Kim D, Xu Y (2003) Protein threading by linear programming. Pac Symp Biocomput .

154. Peng J, Xu J (2011) RaptorX: exploiting structure information for protein alignment by statistical inference. Proteins 79 Suppl 10: 161-171.

155. Källberg M, Wang H, Wang S, Peng J, Wang Z, et al. (2012) Template-based protein structure modeling using the RaptorX web server. Nat Protoc 7: 1511-1522.

156. http://raptorx.uchicago.edu

157. Bonneau R, Baker D (2001) Ab initio protein structure prediction: progress and prospects. Annu Rev Biophys Biomol Struct 30: 173-189.

158. Moult J, Hubbard T, Fidelis K, Pedersen JT (1999) Critical assessment of methods of protein structure prediction (CASP): round III. Proteins 3: 2–6.

159. Skolnick J, Kolinski A (2002) A unified approach to the prediction of protein structure and function. In: Richard A Friesner (Ed.). Advances in Chemical Physics, Computational Methods for Protein Folding (Volume 120) John Wiley & Sons. Inc., NY, USA.]

160. Floudas CA, Fung HK, McAllister SR, Monnigmann M, Rajgaria R (2006) Advances in Protein Structure Prediction and De Novo Protein Design A Review. Chemical Engineering Science 61: 966-988.

161. Zhang Y, Skolnick J (2004) Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. Biophys J 87: 2647-2655.

162. Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res 35: 3375-3382.

163. Zhang Y, Skolnick J (2004) Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci U S A 101: 7594-7599.

164. Karplus K, Barrett C, Hughey R (1998) Hidden Markov models for detecting remote protein homologies. Bioinformatics 14: 846-856.

165. Zhang Y, Kolinski A, Skolnick J (2003) TOUCHSTONE II: a new approach to ab initio protein structure prediction. Biophys J 85: 1145-1164.

166. Zhang Y, Kihara D, Skolnick J (2002) Local energy landscape flattening: parallel hyperbolic Monte Carlo sampling of protein folding. Proteins 48: 192-201.

167. Zhang Y, Skolnick J (2004) SPICKER: a clustering approach to identify near-native protein folds. J Comput Chem 25: 865-871.

168. Berenger F, Zhou Y, Shrestha R, Zhang KY (2011) Entropy-accelerated exact clustering of protein decoys. Bioinformatics 27: 939-945.

169. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33: 2302-2309.

170. Feig M, Rotkiewicz P, Kolinski A, Skolnick J, Brooks CL 3rd (2000) Accurate reconstruction of all-atom protein representations from side-chain-based low-resolution models. Proteins 41: 86-97.

171. Canutescu AA, Shelenkov AA, Dunbrack RL Jr (2003) A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 12: 2001-2014.

172. Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80: 1715-1735.

173. http://zhanglab.ccmb.med.umich.edu/PSSpred

174. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, et al. (2012) PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res 40: W288-293.

175. Maupetit J, Tuffery P, Derreumaux P (2007) A coarse-grained protein force field for folding and structure prediction. Proteins 69: 394-408.

176. Tuffery P, Guyon F, Derreumaux P (2005) Improved greedy algorithm for protein structure reconstruction. J Comput Chem 26: 506-513.

177. Moult J, Pedersen JT, Judson R, Fidelis K (1995) A large-scale experiment to assess protein structure prediction methods. Proteins 23: ii-v.

178. Tramontano A, Cozzetto D, Giorgetti A, Raimondo D (2008) The assessment of methods for protein structure prediction. Methods Mol Biol 413: 43-57.

179. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, et al. (2009) Evaluation of template-based models in CASP8 with standard measures. Proteins 77 Suppl 9: 18-28.

180. Zemla A (2003) LGA: A method for finding 3D similarities in protein structures. Nucleic Acids Res 31: 3370-3374.

181. Ben-David M, Noivirt-Brik O, Paz A, Prilusky J, Sussman JL, et al. (2009) Assessment of CASP8 structure predictions for template free targets. Proteins 77 Suppl 9: 50-65.

182. http://predictioncenter.org/casp10/index.cgi

183. Zhou P, Tian F, Shang Z (2009) 2D depiction of nonbonding interactions for protein complexes. J Comput Chem 30: 940-951.

184. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25: 1605-1612.

185. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 38: 305-320.

186. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 15: 411-428.

187. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14: 33-38, 27-8.

188. Nelson M, Humphrey W, Gursoy A, Dalke A, Kal L, et al. (1995) MDScope-A visual computing environment for structural biology. Computer Physics Communications 91: 111-34. [http://journals.ohiolink.edu/ejc/article.cgi?issn=00104655&issue=v91i1-3&article=111_mavcefsb]

189. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26: 1781-1802.

Advances in Protein Chemistry

Chapter: Protein characterization using modern biophysical techniques

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published August, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Advances in Protein Chemistry

Chapter: Protein characterization using modern biophysical techniques

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published August, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Protein Characterization using Modern Biophysical Techniques

Shuja Shafi Malik[1]* and Tripti Shrivastava[2]

[1]Department of Biochemistry and Molecular Biology, School of Medicine University of Maryland, USA

[2]Translational Health Science and Technology Institute, India

***Corresponding author:** Shuja Shafi Malik, Department of Biochemistry and Molecular Biology, School of Medicine University of Maryland, Baltimore, USA; E-mail: shujasmalik@gmail.com

Introduction

Proteins are an important class of macromolecules in living systems that form observable outcome of genetic inheritance or to say are manifestation of genome. They function in different forms, at varied levels and in diverse ways like structural components of cells and cellular organelles, catalyze biochemical reactions in enzymatic pathways, serve as signaling molecules, receive and respond to stimulus or function as one of the most important components of immune system in form of antibodies. It is through *proteins* that life expresses itself and this importance of *proteins* in maintaining life becomes more obvious in conditions of disease, disorder or unfavorable living conditions. Even a small change in protein sequence or structure can have detrimental outcome, like sickle cell anemia, which is caused by single amino acid substitution of glutamic acid by valine. It is by changing or shuffling composition of their *proteins* that certain living systems circumvent hostile atmospheres to make their survival possible and sustainable. The human immunodeficiency virus (HIV) is one such pathogen that evades host immune system by mutating its own envelope protein. Opportunistic pathogens like *Mycobacterium tuberculosis* survive within hosts for infinitely longer durations of their life by switching over to alternate modes of metabolism i.e. use different set of *proteins* other than their regular life-cycle *proteins* that help them survive in these non-favorable environments. In short it is the protein molecules that act as harbinger of life. Berzelius justifiably called these molecules *proteins*. The term 'protein' has its origin Greek word '*proteois*' which means 'primary' or 'standing in front' [1].

Looking at important roles *proteins* play in living system it is but natural for human interest to delve into exploring structure, function, working, mechanisms and other aspects related to these biological entities. Besides these reasons, studying *proteins* becomes important for fact that in almost all cases of disease and disorder they are targets for different therapies and interventions. Therefore methods and approaches for studying *proteins* have kept on evolving both in terms of multitude of techniques applied and in terms of bettering approaches within individual techniques. *Proteins* can be characterized by exploiting diverse structural, biochemical, electromagnetic, spectroscopic and thermodynamic properties that are bestowed on them by virtue of their composition. But in spite of this diversity the fundamental biochemical composition and rules governing protein architecture are same. Isolation of *proteins* and their subsequent functional and structural characterization utilizes cumulative knowledge from basic common composition and the diversity accorded by that composition. This chapter is discussed in three subsections divided on basis of biochemical and biophysical properties of *proteins* viz. *mass and size, electromagnetic properties and thermodynamic characteristics*.

Methods Based on Mass and Size

Proteins like other molecular entities have mass and size and information about these fundamental characteristics holds an important step in protein characterization. *Proteins* are polymers formed by the association of fundamental structural unit, the amino acid. Simplest of *proteins* in terms of composition and oligomeric nature are monomers. There are *proteins* which are formed by association of identical or non-identical monomeric subunits giving rise to formation of what are called homomeric or heteromeric oligomers. Likewise oligomeric *proteins* differ in the number of subunits from a simple dimeric state to more than decameric composition of assemblies like chromatin remodelers and proteasomes. While as mass is directly proportional to the number of amino acids in a protein, its size is determined by other factors as well, like how amino-acids in a monomeric protein interact or in what manner subunits in multi-subunit protein cooperate determines the shapes *proteins* take like globular, fibrous etc. Therefore knowledge about mass and size of *proteins* holds an important key into elucidating information about *proteins* and this information can be exploited further in studying and understanding their different structural and functional characteristics.

Mass Spectrometry

Mass spectrometry is one of the highly recognized techniques used in eliciting information about molecular masses of

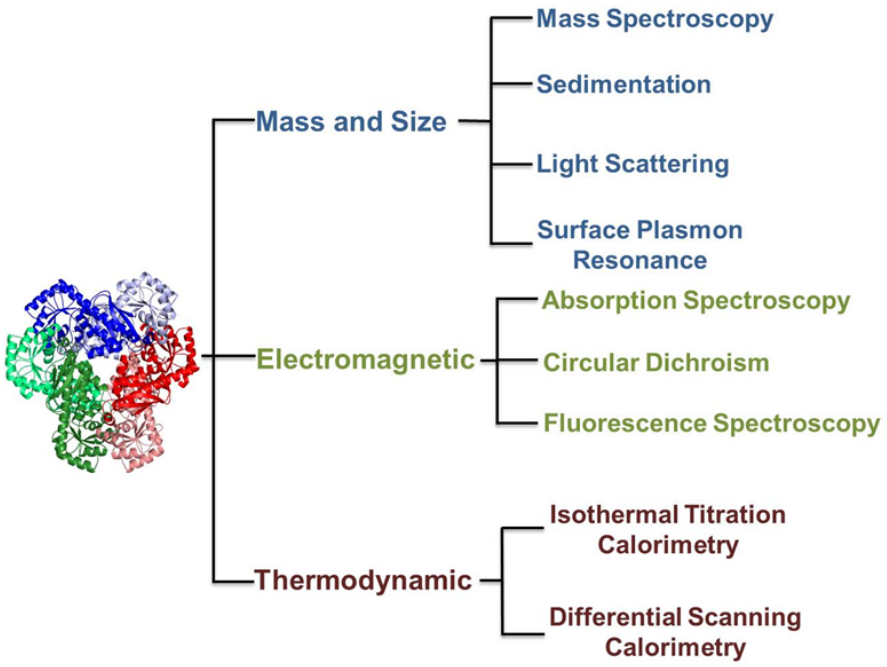


Figure 1: Biochemical and biophysical properties of *proteins* can form basis for their structural and functional characterization.

different molecular species. Precisely it involves separation of molecules according to mass/size ratio of ionized species in an electric field. Mass spectrometry instruments are fundamentally composed of three components: an ionization source that converts particles into ions, a mass analyzer that sorts ions according to *m/z* and an ion detector that measures *m/z* [2]. In these instruments positive ions of analyte are generated in short source region in presence of an electrical field *E*, which imparts kinetic energy K= ZeEs to each ion. In this equation Ze is the charge, *E* the electrical field and *s* the length of source region. The ions generated in this manner into the field-free drift region of length *D* have the same kinetic energy regardless of their size. The velocity with which these ionized particles move is defined by $K = mv^2/2$. Equating the above two equations leads to ZeEs= mv2/2 from which velocity of the particle can be deduced as

$$\mathbf{v} = \left(\frac{2\mathbf{ZeEs}}{\mathbf{m}}\right)^{1/2} \quad (1)$$

The time that each ion will take to traverse the drift region is given by:

$$\mathbf{t} = \frac{\mathbf{D}}{\mathbf{v}} = \left(\frac{\mathbf{m}}{2\mathbf{ZeEs}}\right)^{1/2} \mathbf{D} \quad (2)$$

Combination of equations (1) and (2) allows expressing mass/charge ratio in terms of t/D:

$$\frac{\mathbf{m}}{\mathbf{Z}} = 2\mathbf{ZeEs}\left(\frac{\mathbf{t}}{\mathbf{D}}\right)^2 \quad (3)$$

Based on method of ionization used, mass spectroscopy can be categorized into two types: Matrix-Assisted Laser Desorption-Ionization (MALDI) and Electrospray Ionization (ESI). Briefly, in MALDI, protein containing samples are embedded into specific matrix molecules that absorb the ionization laser beam and transfer energy to analyte. On the other hand in ESI analyte samples are directly injected into the ionizing chamber that converts peptides into smaller ions. Ionized analytes in both methods are directed via a mass analyzer towards a detector that generates MS spectra with each peak of the spectra representing a characteristic m/z ratio.

Application: Role of mass spectrometry in protein characterization has been far beyond merely identifying masses and forms a very formidable technique in 'proteomics'. Thus role for mass spectroscopy in protein characterization can be envisioned in protein quantification, protein profiling, protein interaction analysis and studying modifications like post-translation modifications [3,4]. Protein mass estimation involves by mass spectrometry involves use of single-stage mass spectrometers that act like balances [5]. Proteome analysis on other hand involves use of tandem mass spectrometry (MS/MS) which is accompanied by second stage fragmentation of specific ions after their mass has been determined [5]. MALDI is usually coupled to time of flight (TOF) analyzers that directly measure mass of intact peptides whereas ESI is coupled to ion traps and triple quadrupole analyzers and are used to generate fragment ion spectra (collision-induced (CID) spectra) of selected precursor ions [3,6,7].

Studying *proteins* through mass spectroscopy is a multi-step process. *First step* involves isolation of *proteins* from their biological sources to be followed by digestion and fractionation. While as peptide- mass mapping by MALDI-TOF is useful in peptide identification (peptide mass fingerprinting-PMF), peptide sequencing can be accomplished by ESI-MS/MS [3,8]. There are two approaches that can be employed during protein sequencing. The *top-down approach* does not involve enzymatic digestion of protein sample but relies on transfer of intact protein to gas phase. The main advantage

of this approach is 100% sequence coverage of protein and improved detection of post-translational modifications [9,10]. It is disadvantaged for the reason it requires high field magnetic fields. The most popular and widely used approach for identifying *proteins* and determining details of their sequence and posttranslational modifications is the *bottom-up approach* [6,11]. In *bottom-up approach proteins* of interest are digested with proteolytic enzyme like trypsin and then analyzed by mass spectrometry. In *First step* masses of peptides are determined followed by fractionation of these peptide ions for further downstream analysis. This approach is useful for identifying *proteins* because tryptic peptides solubilize and separate readily than the parent *proteins*. The disadvantage with this approach is that only a small fraction of tryptic peptides are normally detected and as a result some information gets deleted that would be crucial in construction of what are called 'fragmentation ladders'. Whatever the method or methods utilized to generate the data, a large set of data is generated that needs to be further analyzed to derive meaningful information. *Second step* of mass spectrometry based protein characterization is data analysis and interpretation. The *First step* here is deducing amino acid sequence from large datasets generated to be followed by peptide identity. Although many software tools exist for such analysis nevertheless they are deemed to be in need of more accuracy and consistency to reduce results from data redundancy. Of late there has been an emphasis on integrating information from bioinformatics with results that come from proteomics experiments. Three major protein databases SWISS-PROT, TrEMBL and NCBI have been successful in achieving the goals of ability to store, allow searching and retrieving information generated from proteomics [12]. A statistical analysis of the results is an important consideration to ensure confidence in the outcomes [7].

Sedimentation

Sedimentation as the term implies is ability of suspended particles to settle during course of motion as a consequence of effect of different operational forces in sedimenting solution. *Proteins* like other macromolecules can also undergo sedimentation and this ability of *proteins* to settle in solution has been exploited in their characterization through use of analytical ultracentrifugation (AUC).

In analytical ultracentrifugation protein solutions to be studied are subjected to a high gravitational field and resulting changes in concentration distribution are monitored in real time using various optical methods. The optical systems that are currently available for analytical ultracentrifuges include absorbance, fluorescence and interference [13,14]. Although analytical ultracentrifugation places few restrictions on the sample and nature of solvent, there are few fundamental requirements; (i) That sample has a differentiable or distinguishing optical property, (ii) It sediments or floats at an gravitational field that is achievable experimentally and (iii) that it is chemically compatible with the sample cell [13]. The molecular weights that are suitable for AUC vary from between hundred Daltons like peptides, oligosaccharides to million Daltons like viruses and cellular organelles.

Sedimentation during ultracentrifugation can be considered as an outcome of three forces [2,13-16]. The *centrifugal force* that a protein molecule experiences because of spinning is $M_p\omega^2 r$, where M_p is mass of protein, ω is rotor speed in radians/sec ($\omega=2\pi$*rpm/60) and r is the distance from centre of rotor. A *counter force* $M_s\omega^2 r$ exerted on protein molecule is generated by mass M_s of the solvent displaced as the protein molecule sediments. The mass of solvent that is displaced equals Mp * $\overline{v}$ * ρ, where $\overline{v}$ is partial specific volume (in cm^3/g) of the particle and ρ (in g/cm^3) is density of the solvent. Therefore the effective buoyant mass M_b of protein molecule equals $M_p(1-\overline{v}\rho)$. The net force $(M_p-M_s)\omega^2 r$ or $M_p(1-\overline{v}\rho)\omega^2 r$ contributes to the overall viscous drag of protein molecule undergoing sedimentation. This gets balanced by a *frictional force*. This frictional force is represented by fv, where f is the frictional coefficient and v is the velocity. A net outcome will be molecule moving at a velocity that is bare enough to make the total force equivalent to zero.

$$M_p(1-\overline{v}\rho)\,\omega^2 r - fv = 0 \quad (4)$$

Multiplying equation 4 by Avogadro's number (NA) to put entities on molar basis and rearranging to place molecular parameters on one side of equation and experimentally measured ones on other we get

$$\mathbf{NA}*\frac{\mathbf{Mp}(1-\overline{v}\rho)}{f}=\frac{\mathbf{v}}{\omega^2\mathbf{r}}=\mathbf{s} \quad (5)$$

This velocity divided by centrifugal field strength is called *sedimentation coefficient*, *s*. Units of *s* are 'second' and values of 10^{-13} are commonly encountered, the quantity of $1*10^{-13}$ sec has been called 1 Svedberg. Svedberg is named after T. Svedberg a pioneer in sedimentation analysis. Sedimentation coefficient is directly proportional to molecular weight and inversely proportional to frictional coefficient and is experimentally measurable as ratio of velocity to field strength. It can be presumed that sedimentation coefficient will increase with increasing molecular weight; however it also depends on friction *f*, which in turn depends on size, shape and hydration of a protein molecule.

Application: As explained above analytical ultracentrifugation (AUC) relies on properties of mass and fundamental laws of gravitation, it has therefore diverse applicability. Also during analytical centrifugation protein samples can be characterized in their native state and under physiologically and biologically relevant solution conditions. AUC can deliver information about two aspects of solution behavior; hydrodynamic and thermodynamic [2,13-15,17]. Information about hydrodynamic properties like size and shape of protein molecules is deduced from *sedimentation velocity* while as information about thermodynamic properties like molar mass, stoichiometry and association constant comes from *sedimentation equilibrium*.

Sedimentation equilibrium is a more effective method for determination of molecular mass as well as one of the effective methods for characterization of macromolecular interactions. This is for reasons that *sedimentation equilibrium* estimations are not dependent upon macromolecular shape and the reaction kinetics are not part of data analysis [15]. In *sedimentation*

equilibrium an equilibrium concentration distribution of protein macromolecules is obtained where the flux of sedimenting protein molecules is balanced by flux of their diffusion. At this stage the chemical potential of solution is constant, resulting in a concentration gradient that is defined by equilibrium concentration gradient c(r) [13,15]. The study of interacting systems by *sedimentation equilibrium* is made possible by this fact that there is chemical equilibrium at *sedimentation equilibrium* and therefore different components in system like protein and a ligand are related by laws of mass action [15]. *Sedimentation equilibrium* experiments are usually conducted at low rotor speeds so that a balance is created between flux of sedimenting molecules and their diffusion. Extensive use of AUC in study of interactions of *proteins* with nucleic acids [18-21], ligands [22-24] and receptors [22-24] has been made for studying both affinity as well as stoichiometry. AUC has also been employed in studying self-association phenomenon in various regulatory *proteins* and receptors [25-29].

Surface plasmon resonance

Surface Plasmon Resonance (SPR) is an optical technique that helps in detecting changes in refractive indices coming from mass transfer on SPR active surfaces. Reliance on the intrinsic optical properties of protein thus makes SPR a label-free biosensing technology. A typical biosensing experiment involves studying interaction between a ligand in solution (called 'analyte' in SPR terminology) with an immobilized interacting partner (called 'ligand') on an SPR active surface [30,31]. An SPR active surface usually is composed of an SPR-active metal coated glass slide that forms one wall of a thin flow cell, the refractive index changes in which are induced by introduction of analyte solution. SPR occurs when light is shone through the glass slide and onto the metal surface at angles and wavelengths near the so-called "surface plasmon resonance" condition. The incident light couples to oscillations of conducting electrons, the plasmons, at the metal surface and these oscillations in turn create an electromagnetic field commonly referred to as 'evanescent wave' [30,31]. This wave extends from the metal surface into the sample solution on the back side of the surface relative to incident light, i.e. the non-illuminated surface. The intensity of reflected light decreases at sharply defined angle on occurrence of resonance; this angle is called SPR angle and is dependent on refractive index, penetrable by evanescent wave close to metal surface. When other parameters like wavelength of light source or metal of the film are kept constant, the SPR angle shifts are dependent only on changes in refractive index of thin layer adjacent to metal surface. A gradual increase of material resulting from biomolecular interaction at the surface will cause a successive increase of SPR angle and this can be monitored in real time [31-33]. The SPR angle shifts obtained from different protein solutions have been found to be linear within a wide range of surface concentration. In SPR instruments output is given in terms of resonance signal measured in resonance units (RU); 1000 RU corresponds to 0.1° shift in SPR angle and for an average protein this is equivalent to surface concentration change of around 1ng/mm^2 [30,34].

SPR Biosensor chips and Instrumentation: The most common metal in sensor chips consists of gold and the commercially available chips are made up of glass substrate on to which a 50 nm thick gold film is deposited. A monolayer of a long-chain hydroxyalkanethiol that covers the gold film serves as a barrier to prevent analyte from coming in contact with gold [30,31,34]. This also serves as attachment point for carboxymethylated dextran chains that create a hydrophilic surface to which *proteins* are covalently coupled. Interacting partners of SPR active surfaces run through a system of flow channels which are formed when a microfluidic cartridge comes in contact with sensor surface [30,31,33]. Although a significantly diverse number of commercial biosensor instruments are available, the BIAcore systems are the ones that have been very frequently used. The recent versions of BIAcore instrumentation like BIAcore T100, BIAcore A100 and BIAcore X100 also offer thorough method and evaluation software [30].

Application: The major and most important application of SPR has been in analyzing interactions in biomolecules. With huge volume of data coming from various genome sequencing projects and emphasis being on defining functions of annotated genes and hypothetical *proteins*, SPR has proved to be a robust technique in achieving the same.

As mentioned above a typical SPR experiment involves immobilization of one partner from the interacting pair with other being injected across the surface over it. In interaction experiments involving DNA, it is the DNA molecule that is preferably immobilized for reasons it is conformationally flexible as compared to protein hence doesn't affect outcome of the interaction [30]. The DNA molecule to be immobilized is tagged with some molecule through which it can be immobilized on sensor chip. For example a biotin labeled DNA molecule can be immobilized on a sensor chip which is coupled with streptavidin [30]. In case of protein-protein interactions one of the interacting *proteins* can be immobilized by linking to an affinity tag like a poly-histidine tag or a whole protein like GST or MBP [34-37]. The tagged *proteins* can be immobilized on a sensor chip that is coupled to antibody specific to the tag on protein.

The results of interaction analyses experiments are represented in form of a 'sensogram' i.e. a plot of response units (RU) versus time (s). Data obtained is usually analyzed by using software supplied by manufacturer and data analysis is usually performed to elicit information about binding constants and kinetic parameters. To extract correct kinetic parameters from sensogram, the curve must be analyzed in terms of a defined mathematical model [30].

For a simplest case the protein that is injected across the surface after an infinite time arrives at an association equilibrium producing a signal R_{eq}, therefore resonance signal R at time t during this process following injection at t=0 when R=R_0, should obey the expression:

$$\mathbf{R(t) = R_0 + (R_{eq} - R_0)[1 - e^{(-k_{obs}t)}]} \quad (6)$$

The dissociation of bound protein can be expressed by:

$$\mathbf{R(t) = R_0 + (R_{eq} - R_0)[1 - e^{(-k_{off}t)}]} \quad (7)$$

assuming there is complete dissociation of analyte from immobilized ligand. The observed reaction rate for the interaction is given by:

$$k_{obs}=k_{ass}[P]+k_{diss} \quad (8)$$

Thus a linear relationship exists between kobs and total protein concentration [P]. kobs can be directly calculated from equation 6 and by linear regression analysis of dependence of kobs on [P] using equation 8, kass and kdiss can be calculated. For a simple case dissociation constant Kd can be estimated from:

$$k_{diss}=k_{ass}K_d \quad (9)$$

Thus equilibrium dissociation constant (K_d) can be estimated from ratio of *off* and *on* rates. SPR has proved to be a powerful and quantitative tool is studying protein-protein, protein-DNA, protein-ligand and protein-membrane interactions [30,32,38-41]. It provides an advantage of a label-free biochemical assay that helps in identifying interactions, quantifying equilibrium and kinetics constants and underlying energetics. In addition to its application in biological research SPR has been utilized in pharmaceutical research, medical diagnostics, environmental monitoring and food safety and security [41-43].

Light Scattering

Proteins as we know are polypeptide chains with well-defined three dimensional structures. The structure of a protein molecule is also affected by its interactions with its 'surroundings' usually a buffer in which protein is placed. Changes in these buffer or solution conditions can lead to changes in molecular structure of protein including change in size [44]. *Proteins* can also undergo conformational changes leading to change of shape and size during course of interaction with ligands. Thus monitoring the changes in size of *proteins* can be a possible way of studying solution behavior and molecular interactions of *proteins*. Light scattering is a strong method that can be useful in monitoring these changes. The data obtained from light scattering studies can be converted into a meaningful manner to deduce information relevant to characterization of structural and functional properties of *proteins* [45].

When light passes through a solution, electric field of light induces an oscillating polarization of electrons in molecules, hence providing a secondary source of light and subsequently scattered light [45]. The amount of light scattered is directly proportional to the product of the weight-average molar mass (M_w) and the macromolecule concentration (c), i.e. LS ~ M_w*c [45,46]. This scattered light can be analyzed in term of frequency shifts, angular distribution, polarization and the intensity [47]. The measurement of intensity of scattered light as function of angular dispersion is known as *Static Light Scattering* [45,48]. Light scattering can also be used to measure the rate of diffusion of protein particles in solution through phenomenon called *Dynamic Light Scattering* [45].

***Static Light Scattering* (SLS):** *Static Light Scattering* (SLS) also known as Classical light scattering or Rayleigh light scattering is a non-invasive technique used for determination of molar masses (molecular weights) and radii of gyration for macromolecules in solution [45]. SLS measures the average intensity of scattered light scattered by macromolecule in a solution of defined concentration in excess of background scattering of solvent alone. This difference is expressed as *Excess Rayleigh ratio R*, which describes the fraction of incident light scattered by the macromolecules per unit volume of solution [48]. This can be exploited in measuring the weight-average molar mass (Mw) of the scattering particles. When size of the scattering species is smaller than the wavelength of the light source by a factor of 20 it can be expected to behave as a point source. In this case, a simplified form of the Debye equation, based on Rayleigh-Gans-Debye theory, can be applied:

$$(K^*c)/R_\theta=1/M_w+2^*A_2^*c \quad (10)$$

where K an optical parameter and is estimated as shown in equation 11, c is concentration of the scattering species in solution, R_θ is the Rayleigh ratio or the excess intensity of scattered light at scattering angle θ and A_2 is the second osmotic virial coefficient of the scattering species. K can be estimated as

$$\mathbf{K}=\frac{4\pi^2[\mathbf{n}_0\left(\mathbf{dn}/\mathbf{dc}\right)]^2}{\mathbf{N_A}*\lambda^4} \quad (11)$$

where N_A is Avogadro's number, dn/dc is the protein's refractive increment, λ is the wavelength of the laser light source, n_0 is the refractive index of the solvent [45,49].

SLS is useful for determining oligomeric nature and composition of a protein; be it a monomer or a higher oligomer and for measuring the masses of aggregates or other non-native species. It also can be used for measuring the stoichiometry of complexes between different *proteins* e.g. receptor-ligand complexes or antibody-antigen complexes [48,50-52]. An important application of SLS has been developed by combining it with size exclusion chromatography (SEC). In case of higher molecular weight macromolecules or extended *proteins* scattering varies significantly with angle. SEC-MALS involves measurement of scattering at additional angle ("multi-angle light scattering" or MALS) in Size exclusion chromatography (SEC) [48]. In SEC-MALS, Size exclusion chromatography (SEC) system is coupled with on-line laser light scattering (LS), refractive index (RI) and ultraviolet (UV) detector. This approach has been used for the molecular weight determination of *proteins* and their complexes, including glyco*proteins* or integral membrane *proteins* solubilized in detergents [49]. However sometimes *protein aggregation* has been reported to cause problems in its application because of plugging and fouling problems in chromatography columns leading to problems in reproducing results in SEC-MALS. For this reason, SEC-MALS is being replaced with 'field flow fractionation' (FFF, FFF-MALS). FFF relies on a combination of field-driven and diffusive transport mechanisms to separate polymers with wide range, allowing molecular weight and size distributions

for virtually all macromolecules to be determined without the need for filtration. Separation in FFF takes place in an open flow channel. A cross-flow is induced perpendicular to the channel forcing the sample components towards accumulation wall establishing concentration gradient. Smaller particles with higher diffusion coefficients achieve equilibrium positions at higher levels hence eluting first, opposite to the order of elution in SEC [49,53].

***Dynamic Light Scattering* (DLS):** *Dynamic Light Scattering* (DLS) also called photon correlation spectroscopy (PCS) or quasi-elastic light scattering (QELS) allows determination of translational diffusion coefficient D, and hydrodynamic radius R_h [45]. DLS detects fluctuations in scattering intensity due to Brownian motion of molecules in solution. When monochromatic light beam such as laser passes through a solution, the Brownian motion of molecules causes a Doppler shift in wavelength of incident light when it hits a moving particle [54]. Studying changes in wavelength can be a useful tool in identifying size of the particle. It is possible to compute the sphere size distribution and to give a description of the particle's motion in the medium by measuring the diffusion coefficient of the particle. The diffusion coefficient of a sphere at infinite dilution (D_0) is related to hydrodynamic radius (R_h) by Stokes-Einstein relationship,

$$\mathbf{D}_0 = \frac{\mathbf{k}_{\mathbf{b}*\mathbf{T}}}{6\pi\eta\mathbf{R}_\mathbf{h}} \tag{12}$$

where k_b is the Boltzmann constant and η is the solvent viscosity at absolute temperature, T. Dynamic scattering is particularly good at sensing the presence of very small amounts of aggregated protein (<0.01% by weight) and studying samples containing a very large range of masses. It can be quite valuable for comparing stability of different formulations, including real-time monitoring of changes at elevated temperatures [49,54].

DLS has been an effective tool in studying *protein aggregation* as well as in monitoring formation of higher order oligomers in solution. Scattering intensity is proportional to sixth power of molecular radius ($I_0 \alpha\ r^6/\lambda^4$) [55], hence making this technique ideal for identifying the presence of trace amounts of aggregates. Presence of even small aggregation in solution results in significant change in mean hydrodynamic size and also in the percentage polydispersity [44,55,56]. Even if the protein moves from monomer to dimer the associated doubling of size of protein results in six times increase in volume, hence very small amount of change in mass causes quite significant changes on size [49]. *Thermal stability* of *proteins* can be assessed by the measuring scattering intensity and consequently size as function of temperature. Under influence of heat protein molecules denature into massive aggregates and this leads to visible increase in light scattering signifying an increase in size [56]. The marked point where both the size and the intensity start to increase significantly is called the melting point (T_m). Melting point is representative of the stability of the protein under the prevalent solution conditions. *Protein crystallization* uses DLS as a diagnostic tool to improve and screen optimal conditions for crystal growth. Optimization of buffer condition, ionic strength, pH, additives increases the solubility of protein sample typically reducing the aggregation and polydispersity increases the possibility of crystallization [57-59].

Methods Based on Electromagnetic Properties

Proteins are charged and compositionally complex molecules with strong and measurable electromagnetic and spectroscopic properties. Spectroscopic properties in *proteins* come from structural centers called 'chromophores' that exist within protein molecules. Different chromophores render different spectroscopic properties to *proteins* and some chromophores contribute to more than one phenomenon. Absorption spectroscopy, circular dichroism and fluorescence spectroscopy are the three major light based phenomena that have been extensively used in studying and characterizing *proteins*.

Absorption Spectroscopy

According to quantum theory various quantized electronic, vibrational and rotational states are available to a molecule. Absorption of light can lead to transition from a lower energy state to a higher energy state can occur on absorption of light. These transitions are characteristic of the molecular composition of the substance under study and therefore can be exploited in their characterization. Backbone of protein consists of an amide/peptide chain with side chains linked to α carbon atom that is present between amides. In addition to chromophore like characteristics from π-π transitions within these amide bonds, aromatic amino acids tryptophan, tyrosine and phenylalanine also contribute to ability of *proteins* to absorb light [2,60-62]. The primary chromophores and their corresponding absorption wavelengths are tabulated in Table 1.

Chromophore	*Abs. Maximum (λ_{max})*
Tryptophan	280nm
Tyrosine	274nm
Phenylalanine	257nm
Disulphide Bond	250-270nm
Peptide Bond	190-230nm

Table 1: Chromophores in proteins.

Application: The role of absorbance spectroscopy in studying *proteins* has been primarily used in estimation of protein concentration. But has not been limited to concentration estimation and in some cases it has been utilized in studying protein-ligand interactions or probing structural properties of *proteins*.

Protein concentration measurement is major use of absorption spectroscopy in protein characterization and can be measured by using Beer-Lambert equation:

$$A=\varepsilon^{*}c^{*}l \quad (13)$$

where constant *c* in equation 13 is the molar concentration of protein, *A* is absorbance A_{280} and *l* is path length in cm and ε, the extinction coefficient is a protein characteristic quantifying the total contribution of different chromophores present in the protein (2). It is estimated as

$$\varepsilon^{280}\,(M^{-1}\,cm^{-1})=\#Trp^{*}5500+\#Tyr^{*}1490+\#Cys^{*}125 \quad (14)$$

Concentration in mg/ml can be estimated as molar concentration divided by molecular weight of protein.

Conformational changes in *proteins* due to structural perturbations or ligand binding can be monitored using absorption spectroscopy by studying changes in chromophore and consequently in spectroscopic properties of protein [63-65]. These changes in chromophore can arise either due to change in the chromophore environment as a consequence of conformational changes or can come from direct chemical interactions of binding molecule with chromophore. These differences can be quantitatively studied by measuring the change in absorbance (Δ*A*) or what is called 'difference spectrum'. Difference spectra are depicted by plotting differences in absorbance of the protein molecule under study in two conditions; native condition and one in which the difference is intended to study.

In addition to utilization of intrinsic spectral properties in studying *proteins*, external spectroscopic probes can also be used to ascertain protein characteristics. An example of one such probe is 5, 5' dithio-*bis*-(2-nitrobenzoic acid) [DTNB] which is used to measure number of free cysteine sulfhydryl groups in a protein [66,67]. DTNB undergoes a significant change in its absorbance spectrum on forming a disulfide bond and this ability to form disulfide bonds with concomitant change in spectral property is exploited in exploring protein composition. Similarly other probes can be used to monitor changes in regions of protein that are in close proximity to probe binding site.

Circular Dichroism

Circular Dichroism (CD) involves studying optical properties, the principles of which are based in molecular asymmetry or chilarity in the protein molecules. Asymmetry or chilarity in protein molecules means their structures and their mirror images are non-superimposable and this asymmetry becomes observable when plane polarized light passes through protein solution. Plane polarized light is composed of two circularly polarized components of equal magnitude; a counter clockwise rotating component called left-handed or L component and clockwise rotating component called right-handed or R component. An optically active solution like those of *proteins* will differently absorb this left and right circularly polarized light [2,68-72]. This difference can be expressed by Beer-Lambert law as:

$$\Delta A=A_{L}-A_{R}=\varepsilon_{L}{}^{*}l^{*}C-\varepsilon_{R}{}^{*}l^{*}C=\Delta\varepsilon^{*}l^{*}C \quad (15)$$

This difference in absorption results in production of elliptically polarized true light with angle ψ and is represented as ellipticity, θ. Data from CD spectroscopy experiments is presented either in terms of ellipticity [θ] (degrees) or differential absorbance ΔA. The mean residue ellipticity ([θ]mrw,λ) at wavelength λ, is given by:

$$[\theta]_{\mathbf{mrw},\lambda}=\mathbf{MRW}^{*}\frac{[\theta]_{\lambda}}{10^{*}\mathbf{d}^{*}\mathbf{c}} \quad (16)$$

where $[\theta]_{\lambda}$ is the observed ellipticity (degrees) at wavelength λ, d is the path length (cm), and c is the concentration (mg/ml). The Mean Residue Weight (MRW) for protein (in reality peptide bond) is calculated from MRW=M/(N-1), where M is the molecular mass of Protein (in Da), and N is the number of amino acids; the number of peptide bonds is N-1. For most amino acids the MRW is 110±5 Da. If molar concentration of protein is known, the molar ellipticity ($[\theta]_{molar,\lambda}$) at wavelength λ is given by:

$$[\theta]_{\mathbf{molar},\lambda}=\frac{100^{*}[\theta]_{\lambda}}{\mathbf{m}^{*}\mathbf{d}} \quad (17)$$

The units of mean residue ellipticity and molar ellipticity are *deg* cm^{2} $dmol^{-1}$. When data is obtained in absorption units, results are expressed in 'molar differential extinction coefficient' Δε. If the observed difference in absorbance at a certain wavelength of a solution of concentration *m* in a cell of path length d (cm) is ΔA, then Δε is given by:

$$\Delta\varepsilon=\frac{\Delta\mathbf{A}}{\mathbf{m}^{*}\mathbf{d}} \quad (18)$$

The units of molar differential extinction coefficient are M^{-1} cm^{-1}. There is a simple numerical relationship between $[\theta]_{mrw}$ and Δε namely $[\theta]_{mrw}=3298\,\Delta\varepsilon$.

Application: The major center of asymmetry present in *proteins* is the amide group formed by interaction between two amino acids. Further contribution to the optical activity of *proteins* comes from their ability to form well defined super asymmetric secondary structural elements like α-helices, β-sheets and β-turns. These secondary structural elements have distinctive CD spectra which in turn contribute to characteristic CD spectra of individual *proteins*, depending on the relative contribution of structural elements towards overall structure of the protein. Consequently CD spectroscopy has been used as a tool to study structural composition of *proteins*. CD spectroscopy has found all the more important role in studying changes in protein structure and conformation in response to ligand binding or in response to presence of structural perturbants like temperature and denaturing agents.

Secondary structure composition of *proteins* by CD spectroscopy has been studied exploiting the asymmetry generated by peptide bond. This is accomplished by utilizing the data recorded in CD spectra of *proteins* below 250 nm (far UV

region) for estimating fraction or percentage of different secondary structural elements in *proteins* [69,71-73]. The data obtained can be analyzed through different algorithms that are based on the CD spectra of *proteins* of various fold types that have already been solved by X-ray crystallography. Some free programs that are widely used include DICHROWEB [74,75] and K2D2 [76].

Tertiary structure information for *proteins* using CD spectroscopy is ascertained by analyzing CD spectra obtained between wavelengths 260-320nm [68]. Spectrum in this region is a contribution of aromatic amino acids. The characteristics of near UV CD spectrum of a protein like shape and magnitude will depend on many factors like type and number of aromatic amino acids, mobility and nature of their environment and their spatial distribution within the protein. There hasn't been much advancement in use of near UV CD spectra to gain significant structural insights, nevertheless in some cases like bovine ribonuclease [77] and human carbonic anhydrase II [78] assigning features of spectrum to particular amino acids by sequential removal of these residues by site-directed mutagenesis has been accomplished. Near UV CD spectrum can also provide important evidence for existence of "molten globule" states in *proteins*, which are characterized among other things by very weak near UV CD signals [79,80].

Ligand binding induced conformational changes in *proteins* are an essential part of their mechanism of action and biological activity and can be visualized through CD spectroscopy [81-83]. These changes can be monitored both as a function of ligand concentration and function of time (in time-resolved CD studies) [73]. Former leads to information about effective optimal ligand concentration while as latter provides information about response time to generate effect of ligand binding.

Structural integrity and folding properties of *proteins* can also be studied using CD spectroscopy. Contribution of different amino acids towards protein structure can be identified by combination of site-directed mutagenesis and CD spectroscopy. Coupling these experiments to chemical and thermal denaturation experiments can lead to information about relative contribution of amino acids. Mechanism of protein folding using CD spectroscopy has been studied by measuring rate of acquisition of secondary and tertiary structure of a denatured protein. Information like this can be ascertained by studying folding events at millisecond or sub millisecond scales using continuous or stopped-flow CD methods [84-87].

Fluorescence Spectroscopy

Absorption and CD spectroscopy involve excitation from ground state to a higher energy state with expected return to ground state. This return to ground state is accompanied by release of the absorbed energy in a form such as heat that is not further informative. However in some cases absorbed energy is released in other forms that can lead to further information about characteristics of material emitting absorbed energy. Fluorescence is one such mechanism usually classified under 'emission spectroscopy' which involves transitions of electrons between different vibrational states in ground and excited states, the return to ground state being accompanied by release of photons. The ratio of the number of photons emitted to the number of photons absorbed known as fluorescence quantum yield gives information about efficiency of fluorescence process [2,88]. Intrinsic fluorescence in *proteins* is a collective contribution from its aromatic amino acids tryptophan, tyrosine and phenylalanine. They differ in absorption (excitation) and emission wavelengths, fluorescence life times and their fluorescence quantum yields [Table 2]. Due to differences in excitation and emission wavelengths there is resonance energy transfer from phenylalanine to tyrosine and from tyrosine to tryptophan, as a result of which the fluorescence spectrum of a protein mostly resembles that of tryptophan [88].

Fluorophore	*Excitation Wavelength*	*Emission Wavelength*	*Lifetime (10^{-9}sec)*	*Quantum yield*
Tryptophan	280nm	348nm	2.6	0.2
Tyrosine	274nm	303nm	3.6	0.14
Phenylalanine	257nm	282nm	6.4	0.04

Table 2: Fluorophores in proteins.

Application: Fluorescence spectroscopy has been extensively exploited in protein characterization. *Proteins* have intrinsic fluorophores which form readily available centers that can provide information about protein structure. Further quantifiable changes in characteristics of these fluorophores can be monitored for changes in protein conformation. In addition *proteins* and other protein interacting molecules like nucleotides and DNA can be covalently linked to a small fluorescent molecule which makes them very useful in studying macromolecular interactions and enzymatic reactions. Some applications have been elaborated below:

Intrinsic fluorescence spectrum of a protein in native conformation is characteristic of that particular protein, although measure of intensity of fluorescence is not much informative. The value of quantum yield is dependent on other environmental factors like temperature, solvent and presence of other chemical species that might contribute to increase or decrease of fluorescence intensity [2,88,89]. Thus information about presence of tryptophan in a specific environment can be deduced from its emission maxima and to some extent intensity also. A solvent exposed tryptophan having interactions with surrounding residues will have a red-shifted emission maximum as opposed to buried tryptophan residue which will have a blue-shifted maximum [2,88,89]. Of particular interest are those tryptophan residues which are in proximity to a ligand binding site of the protein, because they can lead to information about protein topography [89-92]. *Quenchers* like acrylamide and heavy metal derivatives are used as probe to gain information about molecular make up especially nature of charged residues around active sites [92,93]. Similarly quenchers serve as means to monitor the extent of conformational changes as a result of ligand binding.

Protein** unfolding **kinetics has been observed by monitoring changes in emission maximum or fluorescence intensity due to perturbations in protein structure in presence of destabilizing agents [94-96]. Extrinsic fluorophores also have been used in studying protein stability or protein structure and microenvironment. These extrinsic fluorophores either bind covalently to primary amino groups, or specific group likes thiols or non-covalently on basis of charges or hydrophobicity. Ammonium 8-anilino-1-naphthalenesulfonate or ANS is one such probe that binds to hydrophobic regions of a protein which is accompanied by blue-shift of the fluorescence maxima from 545nm for free ANS to 470nm for bound [97]. Once bound to protein it can be used as probe to monitor protein conformational or structural changes due to altered solution conditions.

Ligand binding to *proteins* has been studied extensively using fluorescence spectroscopy by studying changes in intrinsic tryptophan fluorescence, changes in extrinsic bound fluorophores or by studying changes in the fluorescence properties of the incoming ligand as a result of protein-ligand interaction. Protein-ligand interaction using fluorescence spectroscopy is performed by titration experiments and the extent to which fluorescence is quenched or enhanced is proportional to formation of the protein-ligand complex. Change in fluorescence is measured as a function of ligand concentration and is expressed in form of what is called 'binding isotherm'. This approach can be used to determine the extent of maximal fluorescence change when protein is fully bound to ligand (i.e. at saturation) and consequently fraction of bound and free protein at any concentration of ligand can be determined. Ligand binding at equilibrium in this fashion can be explained by:

$$\mathbf{K_d} = \frac{[\mathbf{P}][\mathbf{L}]}{[\mathbf{P}.\mathbf{L}]} \tag{19}$$

where K_d is the apparent dissociation constant, [P] is the concentration of the protein, [P.L] is the concentration of the protein-ligand complex and [L] is the concentration of unbound ligand. The K_d values can be determined from non-linear least squares (NLLS) regression analysis of titration data using different variants of the above equation. There are many commercially available software packages like GraphPad Prism, GraFit, KaleidaGraph that can be used for treatment of the binding data obtained in fluorescence experiments. These programs have preloaded equations and models that take into account different possibilities that can exist during and after ligand binding to protein; the two important factors being the number of potential binding sites and cooperativity of binding. These software packages also provide options to incorporate equations for models that are not already present and then use them for data fitting and analysis. Few examples of application in studying protein ligand interactions can be found in references [89-92] and [98-100]. Another approach for studying protein-ligand interactions using fluorescence is Macromolecular Competition Titration (MCT) method which is also a thermodynamically rigorous method. MCT is also useful in cases where formation of the complexes is not accompanied by any adequate spectroscopic signal change [101]. It involves quantitative titrations of the reference fluorescent ligand into a protein solution in presence of a fixed concentration (final experiment needs to be conducted over range of concentrations) of a non-fluorescent ligand whose interaction parameters are to be determined. MCT allows determination of total average degree of binding and the free ligand concentrations, over a large degree of binding range, and construction of a model-independent, thermodynamic binding isotherm.

Fluorescence anisotropy (FA) can be employed in such cases in which intensity changes neither in intrinsic fluorescence nor the fluorophore fluorescence are good enough as a parameter of binding data [99,102-105]. To have a certain value of initial anisotropy the fluorophore needs to be attached to a molecule larger is size than itself and for this reason anisotropy based experiments are mostly used in studying *protein-DNA interactions*. A certain disadvantage of fluorescence anisotropy is that it can be used to measure binding constants if molecular size of protein-ligand complex is significantly different from free fluorescing component. Another important consideration while using fluorescence anisotropy is that it relies on an extrinsic fluorophore. This dependence on extrinsic fluorophore is for multiple reasons. First to maximize difference between bound and free states the fluorophore needs to be attached to smaller partner. Second extrinsic fluorophores like ones based on fluorescein or rhodamine derivatives and attached at multiple places to ligand (like on both ends of DNA) have a large extinction coefficient which makes possible conducting experiments at nanomolar range concentrations [99]. In addition to anisotropy, fluorescence polarization can also be used to study protein-ligand interactions [99,105].

Fluorescence resonance energy transfer (FRET) is a fluorescence based phenomenon used to study biological interactions at short distances (1-10 nm). Both interacting species need to have a fluorophore and emission spectrum of donor molecule should overlap with excitation spectrum of acceptor molecule. The extent of overlap between two spectra is referred to as spectral overlap integral (J). Transfer of energy happens through non-radiative dipole-dipole coupling hence donor and acceptor transition dipole orientations must be approximately parallel. Forster has demonstrated that the efficiency of process (E) depends on inverse sixth-distance between donor and acceptor:

$$E = \frac{R_0^6}{R_0^6 + r^6} \tag{20}$$

where R_0 is the Forster distance at which energy transfer is 50% efficient and r is the actual distance between donor and acceptor. For studies involving biological systems distances between 2-9 nm are useful. FRET has been employed in biological systems to usually study protein-DNA and protein-protein interactions. In studying *protein-DNA interactions* the two fluorophores are on DNA molecule at distances where in unbound form they don't involve in energy transfer through FRET. Protein binding to DNA induces changes in the structure of DNA like unwinding or bending that brings the two fluorophores close increasing the FRET efficiency. Increase in FRET efficiency creates changes in the fluorophore of the acceptor molecule which can be monitored as a function of the binding protein molecule that induces this change or can be monitored over time [99,106-108]. FRET *in protein-protein interactions* has been utilized for detecting interactions

between *proteins* in living cells. It is based on fluorescence lifetime imaging (FLIM) which involves measuring lifetime rather than intensity of the donor alone and in presence of acceptor [108]. As expected, to carry out FLIM-FRET experiments each protein in an interaction partnership should have a fluorophore. Different combinations of fluorophores that have been commonly used include EGFP (enhanced-green fluorescent protein) as donor and mCherry (monomeric Cherry red fluorescent variant) as acceptor, CFP (cyan fluorescent protein) as donor and YFP (yellow fluorescent protein) as acceptor [106,109]. In addition to studying intermolecular interactions FRET has been used to study intramolecular phenomenon in *proteins* like protease cleavage, calcium signaling and phosphorylation [106].

Fluorescence spectroscopy in addition to above stated applications in protein characterization has also been used in studying DNA unwinding properties of helicase and other molecular motors like chromatin remodelers. In these experiments usually the changes in fluorescence intensity or anisotropy as a result of local perturbations in DNA structure or gross structural changes in DNA like DNA unwinding are measured. These changes in return are an outcome of the enzymatic activity DNA unwinding enzymes like helicases [110,111].

Methods Based on Thermodynamic Properties

Proteins are bulky molecules formed through innumerable covalent and non-covalent interactions among constituent amino acids. They also interact with solvent molecules in which they are placed. In absence of extraneous factors a chemical and thermodynamic equilibrium is established in such a system [112,113]. Changes in this equilibrium are introduced by presence of factors like other molecular species which can interact with protein as well as with solvent. Interaction between two biomolecules when viewed from atomic level is very complex and thermodynamics of these interactions are determined by number and types of bonds that make up this interaction. These interactions involve formation or breakage of many individual non-covalent bonds like hydrogen bonds, van der Waals interactions and hydrophobic interactions between interacting molecules as well as between interacting molecules and solvent [112-115]. Majority of these events happen around binding site of the protein molecule during which interacting partners come together and solvent molecules are released. This change from 'free' to 'bound' or 'complexed' state can be measured as thermodynamic effect of changes in bond formation in form of overall change in Gibb's energy (G) of the system; enthalpy (H) and entropy (S) being contributors to it.

Calorimetry provides an approach that makes it possible to directly measure the thermodynamic parameters associated with a biomolecular interaction. The direct detection of small heat changes accompanying these interactions provides a universal detection method for studying biomolecular interactions since enthalpy change provides the measure of interaction [112,113]. Other techniques like those based on spectroscopy although provide measurements of binding constants but for estimating enthalpy and other thermodynamic parameters indirect approach by using van't Hoff relationship has to be employed [112]. Furthermore in approaches that are based on portioning of two molecular components like fluorescence spectroscopy require respective amounts of free and bound material to be determined at a given concentration of one of the interacting partners [112]. Consequently this has to be done at a range of concentrations which can make them expensive in terms of both time and material.

Isothermal Titration Calorimetry (ITC) and Differential Scanning Calorimetry (DSC) are two techniques based in thermodynamics that have been extensively employed in biophysical characterization of *proteins*. Their application and mathematical basis of it is briefly discussed in the following sections.

Isothermal titration calorimetry

Isothermal Titration Calorimetry (ITC) is a powerful, fast and accurate method for studying binding affinities and thermodynamics in macromolecular interactions under both isothermal and isobaric conditions. An ITC experiment involves titration of one binding partner at a constant temperature into a known amount of other binding partner placed in sample cell of the calorimeter. The instrument consists of two identical cells housed in an isothermal jacket which is cooled to keep cells and their contents at experimental temperature. These two cells are kept at thermal equilibrium ($\Delta T=0$) during the course of experiment and a control reference cell is filled with buffer. The basis of experiment is to measure heat energy per unit time produced or released on binding. The instrument measures electrical power (in units J/s or Watts) required maintaining zero temperature difference between sample cell and reference cell at the designated experiment temperature [114,116]. For a simple 1:1 interaction between macromolecule M placed in calorimetric cell with ligand L injected into the cell,

$$M+L \leftrightarrow ML \tag{21}$$

According to law of mass action:

$$\mathbf{K_A} = \frac{[\mathbf{ML}]}{[\mathbf{M}][\mathbf{L}]} \tag{22}$$

Total concentration of ligand and protein can be estimated from:

$$L_t=[L]+[ML] \tag{23}$$

$$\mathbf{M_t} = [\mathbf{M}]+[\mathbf{ML}] = [\mathbf{PL}]+\frac{[\mathbf{ML}]}{\mathbf{K_A}[\mathbf{L}]} \tag{24}$$

The total heat content Q of the solution in calorimeter cell (of volume V) is given by:

$$Q=nFM_t \Delta HV \quad (25)$$

where M_t is total macromolecule concentration in calorimetric cell, F is the fraction of sites on M occupied by L and n is the number of sites [112,116]. Substituting for the value of M_t from equation 21 in equation 22 leads to estimation of dissociation constant ($K_d=1/K_A$). (More elaborate treatment of equations can be looked in [112,114,116]. Final data is plotted as differential heat evolved for each injection of an aliquot of ligand L into sample in measurement cell versus the molar ratio L_t/M_t. Data fitting is performed using non-linear least squares (NLLS) method and initial guesses are made for K_A, ΔH and n and are varied till best fit of data is obtained [116]. Thus in ITC in a single experiment the values of the dissociation constant (K_d), the stoichiometry (n) and the enthalpy of binding (ΔH) are determined. The free energy and entropy of binding are determined from:

$$\Delta G=-RtlnKA=\Delta H-T\Delta S \quad (26)$$

where ΔG is Gibbs free energy of binding, R is the universal gas constant, T is the temperature in Kelvin, K_A is the association constant, ΔH is enthalpy change (heat) and ΔS is the entropy change on binding.

As ITC is a label-free method that can measure any reaction that results in heat change, it has found wide application in protein characterization like studying protein-DNA [112,114,115], protein-protein [117,118] and protein-small molecule interactions [113,117-120]. ITC happens to be the only technique that can directly measure energetics including Gibbs free energy, enthalpy, entropy and heat capacity changes. Consequently in addition to binding thermodynamics ITC can also be used to study catalytic reactions, conformational rearrangements and molecular dissociations [117,121,122]. An important role for ITC has come up in structure-based drug design and other drug discovery strategies [117-119]. Free energy changes associated with molecular interactions are a result of characteristics of interaction, like burial of apolar surfaces or decrease in exposure of nonpolar surface or may be accompanied by displacement of buried water molecules. Information from insights into thermodynamics of interaction of drug molecules with *proteins* has been helpful in aiding the design of inhibitors and drug like molecules.

Differential scanning calorimetry

While ITC is conducted in presence of constant temperature, Differential Scanning Calorimetry (DSC) measures heat capacity (C_p) and enthalpy of thermally induced transitions in particular conformational transitions in biological molecules [113,114,123-125]. To put it simply, DSC provides a complete thermodynamic profile for *unfolding* energetics of the system [113]. A typical DSC experiment involves monitoring difference of temperature in two cells; a reference cell that contains reaction buffer and the sample cell that contains the reaction mixture which can be protein alone or mixture of protein and an interacting molecule [114,123,126]. Both cells are electrically heated at a known constant temperature. A temperature induced transition typically endothermic in nature is created in both cells but temperature of the sample cell lags behind that of reference cell. An amount of compensatory electric power driven by a feedback mechanism is used to maintain the two cells at same temperature. This amount of power (units Js^{-1}) at temperature T divided by the heating rate (units Ks^{-1}) is the apparent difference in heat capacity between sample cell (C_p^{solv}) and reference cell (C_p^{solv}), ΔC_p (units JK^{-1}). This power difference can be expressed as:

$$\Delta \mathbf{C_p} = \mathbf{C_p^{sol}} - \mathbf{C_p^{solv}} \quad (27)$$

Thermal transition curves in terms of heat capacity are inadequate because they are normalized to mass of substrate (C_p can have units JK^{-1} or $JK^{-1}g^{-1}$). A more meaningful comparison of the thermal transition curves of different substrates or same substrate in different conditions (like ligand binding) can be achieved by normalizing concentrations according to molarity. This new entity is called *molar heat capacity* (units JK-1mol-1) and is obtained by multiplying C_p by the molecular weight (g mol^{-1}).

For simple *proteins* a single heat absorption peak called 'thermogram' is usually observed in the scan. Applying correction for instrument baseline and transition baseline and normalizing concentration, a direct calorimetric measurement of enthalpy (ΔH) and melting temperature (T_m) can be obtained by integration of the thermogram with respect to temperature.

$$\Delta \mathbf{H} = \int_{T_0}^{T} \mathbf{Cp}.\mathbf{dT} \quad (28)$$

In protein characterization DSC has found application in studying equilibrium thermodynamic stability and folding mechanism [123-126]. Protein stability through DSC is observed by estimating *heat capacity increment* that accompanies protein denaturation or in simpler terms *heat capacity of denaturation*. It is positive because heat capacity of unfolded protein is greater than that of native protein [123,124]. Melting temperature (T_m) is a good indicator of protein thermo stability and generally higher the T_m, the more thermodynamically stable protein is. In addition to protein stability studies as mentioned above and in previous section, DSC along with ITC form a formidable combination that have a role in studying protein interactions [113-117,124-129]. The role of DSC has become all the more important when studying *protein-DNA interactions* [114,130,131]. Since both DNA and protein structure significantly depend upon temperature the combined use of ITC with DSC is made with data obtained from DSC experiments used for correcting ITC data.

Summary

Major aim of protein characterization is to arrive at meaningful and empirical information about these biomolecules. This information is a

pre-requisite in understanding how they function and can also be useful in aspects like drug discovery and development. Gathering reliable and interpretable information about *proteins* like any other scientific process involves great deal of painstaking, dedicated and systematic efforts and can't be achieved by simplistic approaches. A proper strategy involves applying methodologies that lead to protein isolation, biochemical and biophysical characterization and finally structural elucidation. No single technique or approach can be said to be fool proof or robust enough to lead to conclusive results therefore usually a combination of techniques is put to use.

References

1. Tanford C, Reynolds J (2003) Nature's Robots: A History of Proteins. Oxford University Press, Oxford, UK.
2. van Holde KE, Johnson WC, Pui Shing Ho (1998) Principles of Physical Biochemistry. Pearson Education Inc., USA.
3. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422: 198-207.
4. Aebersold R, Goodlett DR (2001) Mass spectrometry in proteomics. Chem Rev 101: 269-295.
5. Nesvizhskii AI (2012) Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. Proteomics 12: 1639-1655.
6. Wysocki VH, Resing KA, Zhang Q, Cheng G (2005) Mass spectrometry of peptides and proteins. Methods 35: 211-222.
7. Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. Science 312: 212-217.
8. Ahram M, Petricoin EF (2008) Proteomics Discovery of Disease Biomarkers. Biomark Insights 3: 325-333.
9. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, et al. (2001) Informatics and multiplexing of intact protein identification in bacteria and the archaea. Nat Biotechnol 19: 952-957.
10. Forbes AJ, Patrie SM, Taylor GK, Kim YB, Jiang L, et al. (2004) Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS. Proc Natl Acad Sci USA 101: 2678-2683.
11. Chait BT (2006) Chemistry. Mass spectrometry: bottom-up or top-down? Science 314: 65-66.
12. Johnson RS, Davis MT, Taylor JA, Patterson SD (2005) Informatics for protein identification by mass spectrometry. Methods 35: 223-236.
13. Cole JL, Lary JW, P Moody T, Laue TM (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. Methods Cell Biol 84: 143-179.
14. Howlett GJ, Minton AP, Rivas G (2006) Analytical ultracentrifugation for the study of protein association and assembly. Curr Opin Chem Biol 10: 430-436.
15. Ghirlando R (2011) The analysis of macromolecular interactions by sedimentation equilibrium. Methods 54: 145-156.
16. Padrick SB, Brautigam CA (2011) Evaluating the stoichiometry of macromolecular complexes using multisignal sedimentation velocity. Methods 54: 39-55.
17. Cole JL, Correia JJ, Stafford WF (2011) The use of analytical sedimentation velocity to extract thermodynamic linkage. Biophys Chem 159: 120-128.
18. Nowotny M, Cerritelli SM, Ghirlando R, Gaidamakov SA, Crouch RJ, et al. (2008) Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. EMBO J 27: 1172-1181.
19. Nowotny M, Gaidamakov SA, Ghirlando R, Cerritelli SM, Crouch RJ, et al. (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. Mol Cell 28: 264-276.
20. Leonard JN, Ghirlando R, Askins J, Bell JK, Margulies DH, et al. (2008) The TLR3 signaling complex forms by cooperative receptor dimerization. Proc Natl Acad Sci US A 105: 258-263.
21. Carroll KL, Ghirlando R, Ames JM, Corden JL (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA 13: 361-373.
22. McLellan JS, Zheng X, Hauk G, Ghirlando R, Beachy PA, et al. (2008) The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. Nature 455: 979-983.
23. Gonçalves KA, Borges JC, Silva JC, Papa PF, Bressan GC, et al. (2010) Solution structure of the human signaling protein RACK1. BMC Struct Biol 10: 15.
24. Snyder GA, Ford J, Torabi-Parizi P, Arthos JA, Schuck P, et al. (2005) Characterization of DC-SIGN/R interaction with human immunodeficiency virus type 1 gp120 and ICAM molecules favors the receptor's role as an antigen-capturing rather than an adhesion receptor. J Virol 79: 4589-4598.
25. Gilmore JM, Bieber Urbauer RJ, Minakhin L, Akoyev V, Zolkiewski M, et al. (2010) Determinants of affinity and activity of the anti-sigma factor AsiA. Biochemistry 49: 6143-6154.
26. Ramsey JE, Daugherty MA, Kelm RJ Jr (2007) Hydrodynamic studies on the quaternary structure of recombinant mouse Purbeta. J Biol Chem 282: 1552-1560.
27. Pesiridis GS, Diamond E, Van Duyne GD (2009) Role of pICLn in methylation of Sm proteins by PRMT5. J Biol Chem 284: 21347-21359.
28. Chaudhry C, Plested AJ, Schuck P, Mayer ML (2009) Energetics of glutamate receptor ligand binding domain dimer assembly are modulated by allosteric ions. Proc Natl Acad Sci U S A 106: 12329-12334.
29. Chaudhry C, Weston MC, Schuck P, Rosenmund C, Mayer ML (2009) Stability of ligand-binding domain dimer assembly controls kainate receptor desensitization. EMBO J 28: 1518-1530.
30. Stockley PG, Persson B (2009) Surface plasmon resonance assays of DNA-protein interactions. Methods Mol Biol 543: 653-669.
31. McDonnell JM (2001) Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. Curr Opin Chem Biol 5: 572-577.
32. Boozer C, Kim G, Cong S, Guan H, Londergan T (2006) Looking towards label-free biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon resonance technologies. Curr Opin Biotechnol 17: 400-405.
33. Piliarik M, Vaisocherová H, Homola J (2009) Surface plasmon resonance biosensing. Methods Mol Biol 503: 65-88.
34. Rich RL, Hoth LR, Geoghegan KF, Brown TA, LeMotte PK, et al. (2002) Kinetic analysis of estrogen receptor/ligand interactions. Proc Natl Acad Sci U S A 99: 8562-8567.
35. Wassaf D, Kuang G, Kopacz K, Wu QL, Nguyen Q, et al. (2006) High-throughput affinity ranking of antibodies using surface plasmon resonance microarrays. Anal Biochem 351: 241-253.
36. Phillips KS, Cheng Q (2007) Recent advances in surface plasmon resonance based techniques for bioanalysis. Anal Bioanal Chem 387: 1831-1840.
37. Rich RL, Myszka DG (2000) Survey of the 1999 surface plasmon resonance biosensor literature. J Mol Recognit 13: 388-407.
38. Ritzefeld M, Sewald N (2012) Real-Time Analysis of Specific Protein-DNA Interactions with Surface Plasmon Resonance. J Amino Acids 2012: 816032.

39. Katsamba PS, Park S, Laird-Offringa IA (2002) Kinetic studies of RNA-protein interactions using surface plasmon resonance. Methods 26: 95-104.

40. Haruki M, Noguchi E, Kanaya S, Crouch RJ (1997) Kinetic and stoichiometric analysis for the binding of Escherichia coli ribonuclease HI to RNA-DNA hybrids using surface plasmon resonance. J Biol Chem 272: 22015-22022.

41. Tsoi PY, Yang M (2002) Kinetic study of various binding modes between human DNA polymerase beta and different DNA substrates by surface-plasmon-resonance biosensor. Biochem J 361: 317-325.

42. Persson B, Stenhag K, Nilsson P, Larsson A, Uhlén M, et al. (1997) Analysis of oligonucleotide probe affinities using surface plasmon resonance: a means for mutational scanning. Anal Biochem 246: 34-44.

43. Di Primo C, Lebars I (2007) Determination of refractive index increment ratios for protein-nucleic acid complexes by surface plasmon resonance. Anal Biochem 368: 148-155.

44. Some D (2013) Light-scattering-based analysis of biomolecular interactions. Biophys Rev 5: 147-158.

45. Nobbmann U, Connah M, Fish B, Varley P, Gee C, et al. (2007) Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclo006Eal antibodies. Biotechnol Genet Eng Rev 24: 117-128.

46. Chi-san Wu (2003) Handbook Of Size Exclusion Chromatography And Related Techniques: Revised and Expanded. CRC Press, Florida, USA.

47. Bruce J. Berne, Robert Pecora (2000) Dynamic Light Scattering: With Applications to Chemistry, Biology and Physics. Courier Dover Publications, USA.

48. Daniel Some, Sophia Kenrick (2012) Protein Interactions, Biochemistry, Genetics and Molecular Biology.

49. Hoppe CC, Nguyen LT, Kirsch LE, Wiencek JM (2008) Characterization of seed nuclei in glucagon aggregation using light scattering methods and field-flow fractionation. J Biol Eng 2: 10.

50. Arthur KK, Gabrielson JP, Hawkins N, Anafi D, Wypych J, et al. (2012) In vitro stoichiometry of complexes between the soluble RANK ligand and the monoclonal antibody denosumab. Biochemistry 51: 795-806.

51. Fountoulakis M, Zulauf M, Lustig A, Garotta G (1992) Stoichiometry of interaction between interferon gamma and its receptor. Eur J Biochem 208: 781-787.

52. Mogridge J (2004) Using light scattering to determine the stoichiometry of protein complexes. Methods Mol Biol 261: 113-118.

53. Reschiglian P, Rambaldi DC, Zattoni A (2011) Flow field-flow fractionation with multiangle light scattering detection for the analysis and characterization of functional nanoparticles. Anal Bioanal Chem 399: 197-203.

54. W I Goldburg (1999) Dynamic light scattering. American Journal of Physics 67: 1152-1160.

55. Peter Brimblecombe (1996) Air composition and chemistry. In: Cambridge environmental chemistry series 6. (2ndedn), Cambridge University Press, Cambridge, UK.

56. Shiba K, Niidome T, Katoh E, Xiang H, Han L, et al. (2010) Polydispersity as a parameter for indicating the thermal stability of proteins by dynamic light scattering. Anal Sci 26: 659-663.

57. Borgstahl GE (2007) How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals. Methods Mol Biol 363: 109-129.

58. Wilson WW (2003) Light scattering as a diagnostic for protein crystal growth--a practical approach. J Struct Biol 142: 56-65.

59. Habel JE, Ohren JF, Borgstahl GE (2001) Dynamic light-scattering analysis of full-length human RPA14/32 dimer: purification, crystallization and self-association. Acta Crystallogr D Biol Crystallogr 57: 254-259.

60. Layne E (1957) Spectrophotometric and Turbidimetric Methods for Measuring Proteins. Methods in Enzymology 3: 447-455.

61. Stoscheck CM (1990) Quantitation of Proteins. Methods in Enzymology 182: 50-68.

62. Aitken A, Leamonth MP (2002) Protein Determination by UV Absorption. Springer Protocols: 3-6.

63. Nienhaus K, Nienhaus GU (2005) Probing heme protein-ligand interactions by UV/visible absorption spectroscopy. Methods Mol Biol 305: 215-242.

64. Nienhaus K, Kriegl JM, Nienhaus GU (2004) Structural dynamics in the active site of murine neuroglobin and its effects on ligand binding. J Biol Chem 279: 22944-22952.

65. Giordano D, Boron I, Abbruzzetti S, Van Leuven W, Nicoletti FP, et al. (2012) Biophysical characterisation of neuroglobin of the icefish, a natural knockout for hemoglobin and myoglobin. Comparison with human neuroglobin. PLoS One 7: 44508.

66. Sedlak J, Lindsay RH (1968) Estimation of total, protein-bound, and nonprotein sulfhydryl groups in tissue with Ellman's reagent. Anal Biochem 25: 192-205.

67. Riener CK, Kada G, Gruber HJ (2002) Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine. Anal Bioanal Chem 373: 266-276.

68. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta 1751: 119-139.

69. Martin SR, Schilstra MJ (2008) Circular dichroism and its application to the study of biomolecules. Methods Cell Biol 84: 263-293.

70. Corrêa DHA, Ramos CHI (2009) The use of circular dichroism spectroscopy to study protein folding, form and function. African Journal of Biochemistry Research 3: 164-173.

71. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1: 2876-2890.

72. Dodero VI, Quirolo ZB, Sequeira MA (2011) Biomolecular studies by circular dichroism. Front Biosci (Landmark Ed) 16: 61-73.

73. Greenfield NJ (2006) Analysis of the kinetics of folding of proteins and peptides using circular dichroism. Nat Protoc 1: 2891-2899.

74. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res 32: W668-673.

75. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89: 392-400.

76. Perez-Iratxeta C, Andrade-Navarro MA (2008) K2D2: estimation of protein secondary structure from circular dichroism spectra. BMC Struct Biol 8: 25.

77. Woody AY, Woody RW (2003) Individual tyrosine side-chain contributions to circular dichroism of ribonuclease. Biopolymers 72: 500-513.

78. Freskgård PO, Mårtensson LG, Jonasson P, Jonsson BH, Carlsson U (1994) Assignment of the contribution of the tryptophan residues to the circular dichroism spectrum of human carbonic anhydrase II. Biochemistry 33: 14281-14288.

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

80. Price NE, Price NC, Kelly SM, McDonnell JM (2005) The key role of protein flexibility in modulating IgE interactions. J Biol Chem 280: 2324-2330.

81. Kelly SM, Price NC (2006) Circular dichroism to study protein interactions. Curr Protoc Protein Sci Chapter 20: Unit 20.

82. Rodger A, Marrington R, Roper D, Windsor S (2005) Circular dichroism spectroscopy for the study of protein-ligand interactions. Methods Mol Biol 305: 343-364.

83. Cary PD, Kneale GG (2009) Circular dichroism for the analysis of protein-DNA interactions. Methods Mol Biol 543: 613-624.

84. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc 1: 2527-2535.

85. Zitzewitz JA, Bilsel O, Luo J, Jones BE, Matthews CR (1995) Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. Biochemistry 34: 12812-12819.

86. Meisner WK, Sosnick TR (2004) Fast folding of a helical protein initiated by the collision of unstructured chains. Proc Natl Acad Sci U S A 101: 13478-13482.

87. Stump MR, Gloss LM (2010) Mutational studies uncover non-native structure in the dimeric kinetic intermediate of the H2A-H2B heterodimer. J Mol Biol 401: 518-531.

88. Lakowicz JR (2006) Principles of Fluorescence Spectroscopy. (3rdedn), Springer, Berlin, Germany.

89. Bujalowski W, Klonowska MM (1994) Close proximity of tryptophan residues and ATP-binding site in Escherichia coli primary replicative helicase DnaB protein. Molecular topography of the enzyme. J Biol Chem 269: 31359-31371.

90. Roychowdhury A, Szymanski MR, Jezewska MJ, Bujalowski W (2009) Interactions of the Escherichia coli DnaB-DnaC protein complex with nucleotide cofactors. 1. Allosteric conformational transitions of the complex. Biochemistry 48: 6712-6729.

91. Roychowdhury A, Szymanski MR, Jezewska MJ, Bujalowski W (2009) Escherichia coli DnaB helicase-DnaC protein complex: allosteric effects of the nucleotides on the nucleic acid binding and the kinetic mechanism of NTP hydrolysis. 3. Biochemistry 48: 6747-6763.

92. Malik SS, Luthra A, Srivastava SK, Ramachandran R (2008) Mycobacterium tuberculosis UsfX (Rv3287c) exhibits novel nucleotide binding and hydrolysis properties. Biochem Biophys Res Commun 375: 465-470.

93. Zhou H, Brock J, Casarotto MG, Oakley AJ, Board PG (2011) Novel folding and stability defects cause a deficiency of human glutathione transferase omega 1. J Biol Chem 286: 4271-4279.

94. Manyusa S, Whitford D (1999) Defining folding and unfolding reactions of apocytochrome b5 using equilibrium and kinetic fluorescence measurements. Biochemistry 38: 9533-9540.

95. Tyler RC, Murray NJ, Peterson FC, Volkman BF (2011) Native-state interconversion of a metamorphic protein requires global unfolding. Biochemistry 50: 7077-7079.

96. Tyler RC, Wieting JC, Peterson FC, Volkman BF (2012) Electrostatic optimization of the conformational energy landscape in a metamorphic protein. Biochemistry 51: 9067-9075.

97. Ludwig HC, Pardo FN, Asenjo JL, Maureira MA, Yañez AJ, et al. (2007) Unraveling multistate unfolding of pig kidney fructose-1,6-bisphosphatase using single tryptophan mutants. FEBS J 274: 5337-5349.

98. Picard-Jean F, Bougie I, Bisaillon M (2007) Characterization of the DNA- and dNTP-binding activities of the human cytomegalovirus DNA polymerase catalytic subunit UL54. Biochem J 407: 331-341.

99. Favicchio R, Dragan AI, Kneale GG, Read CM (2009) Fluorescence spectroscopy and anisotropy in the analysis of DNA-protein interactions. Methods Mol Biol 543: 589-611.

100. Kozlov AG, Galletto R, Lohman TM (2012) SSB-DNA binding monitored by fluorescence intensity and anisotropy. Methods Mol Biol 922: 55-83.

101. Bujalowski W, Jezewska MJ (2011) Macromolecular competition titration method accessing thermodynamics of the unmodified macromolecule-ligand interactions through spectroscopic titrations of fluorescent analogs. Methods Enzymology 488: 17-57.

102. Malik SS, Rich E, Viswanathan R, Cairns BR, Fischer CJ (2011) Allosteric interactions of DNA and nucleotides with S. cerevisiae RSC. Biochemistry 50: 7881-7890.

103. Arosio D, Costantini S, Kong Y, Vindigni A (2004) Fluorescence anisotropy studies on the Ku-DNA interaction: anion and cation effects. J Biol Chem 279: 42826-42835.

104. Pogenberg V, Guichou JF, Vivat-Hannah V, Kammerer S, Pérez E, et al. (2005) Characterization of the interaction between retinoic acid receptor/retinoid X receptor (RAR/RXR) heterodimers and transcriptional coactivators through structural and fluorescence anisotropy studies. JBC 280: 1625-33.

105. Rossi AM, Taylor CW (2011) Analysis of protein-ligand interactions by fluorescence polarization. Nat Protoc 6: 365-387.

106. Truong K, Ikura M (2001) The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. Curr Opin Struct Biol 11: 573-578.

107. Heilemann M, Hwang LC, Lymperopoulos K, Kapanidis AN (2009) Single-molecule FRET analysis of protein-DNA complexes. Methods Mol Biol 543: 503-521.

108. Llères D, Swift S, Lamond AI (2007) Detecting protein-protein interactions in vivo with FRET using multiphoton fluorescence lifetime imaging microscopy (FLIM). Curr Protoc Cytom Chapter 12: Unit12.

109. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, et al. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat Biotechnol 22: 1567-1572.

110. Eastlund A, Malik SS, Fischer CJ (2013) Kinetic mechanism of DNA translocation by the RSC molecular motor. Arch Biochem Biophys 532: 73-83.

111. Khaki AR, Field C, Malik S, Niedziela-Majka A, Leavitt SA, et al. (2010) The macroscopic rate of nucleic acid translocation by hepatitis C virus helicase NS3h is dependent on both sugar and base moieties. J Mol Biol 400: 354-378.

112. 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.

113. Garbett NC, Chaires JB (2012) Thermodynamic studies for drug design and screening. Expert Opin Drug Discov 7: 299-314.

114. Crane-Robinson C, Dragan AI, Read CM (2009) Defining the thermodynamics of protein/DNA complexes and their components using micro-calorimetry. Methods Mol Biol 543: 625-651.

115. Read CM, Jelesarov I (2001) Calorimetry of protein-DNA complexes and their components. Methods Mol Biol 148: 511-533.

116. Recht MI, Ryder SP, Williamson JR (2008) Monitoring assembly of ribonucleoprotein complexes by isothermal titration calorimetry. Methods Mol Biol 488: 117-127.

117. Weber PC, Salemme FR (2003) Applications of calorimetric methods to drug discovery and the study of protein interactions. Curr Opin Struct Biol 13: 115-121.

118. Torres FE, Recht MI, Coyle JE, Bruce RH, Williams G (2010) Higher throughput calorimetry: opportunities, approaches and challenges. Curr Opin Struct Biol 20: 598-605.

119. Ladbury JE (2010) Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. Biochem Soc Trans 38: 888-893.

120. Shrivastava T, Dey A, Ramachandran R (2009) Ligand-induced structural transitions, mutational analysis, and 'open' quaternary structure of M. tuberculosis feast/famine regulatory protein (Rv3291c). JMB 392: 1007-1019.

121. Falconer RJ, Penkova A, Jelesarov I, Collins BM (2010) Survey of the year 2008: applications of isothermal titration calorimetry. J Mol Recognit 23: 395-413.

122. BjeliÄ‡ S, Jelesarov I (2008) A survey of the year 2007 literature on applications of isothermal titration calorimetry. J Mol Recognit 21: 289-312.

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

124. Jelesarov I, Bosshard HR (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. J Mol Recognit 12: 3-18.

125. Chiu MH, Prenner EJ (2011) Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. J Pharm Bioallied Sci 3: 39-59.

126. Celej MS, Montich GG, Fidelio GD (2003) Protein stability induced by ligand binding correlates with changes in protein flexibility. Protein Sci 12: 1496-1506.

127. Celej MS, Dassie SA, González M, Bianconi ML, Fidelio GD (2006) Differential scanning calorimetry as a tool to estimate binding parameters in multiligand binding proteins. Anal Biochem 350: 277-284.

128. Bruylants G, Wouters J, Michaux C (2005) Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. Curr Med Chem 12: 2011-2020.

129. Cañadas O, Casals C (2013) Differential scanning calorimetry of protein-lipid interactions. Methods Mol Biol 974: 55-71.

130. Dragan AI, Klass J, Read C, Churchill ME, Crane-Robinson C, et al. (2003) DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution. J Mol Biol 331: 795-813.

131. Dragan AI, Read CM, Makeyeva EN, Milgotina EI, Churchill ME, et al. (2004) DNA binding and bending by HMG boxes: energetic determinants of specificity. J Mol Biol 343: 371-393.

New Peptide Based Therapeutic Approaches

Riyasat Ali[1]*, Richa Rani[2] and Sudhir Kumar[3]

[1]Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 110029, India
2Department of Biotechnology, Indian Institute of Technology, Roorkee 247667, Uttarakhand, India
[3]Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH 45267-0056, USA

***Corresponding author:** Riyasat Ali, Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 110029, India, E-mail: riyasataiims@gmail.com

Abstract

A large number of chemically well defined and synthetic linear as well as branched therapeutics peptides have been developed and used for a large number of therapeutic applications. Peptides and peptide analogues have been developed to mimic naturally occurring molecules like hormones; growth factors etc. along with recently developed several peptide inhibitors. Moreover, large numbers of studies have been performed to establish peptides as an immune prophylactic and immune diagnostic for several infectious diseases like malaria, HIV and plague etc. Synthetic peptide approach have been popularised due to its ease of synthesising large amount of peptides and liberty to modify peptides in order to make them more effective. In the last couple of decades, with the advancement of technology, it has become easy to synthesise longer peptides using peptide conjugation and modification strategies. Moreover, wide range of targeted delivery vehicles of peptides using several bio-compatible polymers have been developed which made peptide drug approach more popular for therapeutic purposes. Targeted and drug release kinetics for optimum drug dispensing have been greatly advanced which increases bio-availability and stability of peptides in the system for longer therapeutic effects.

Introduction

The enormous increase in the cost and time span to develop a conventional drugs led researchers and pharmaceutical companies to develop new cost-effective products based of synthetic peptide strategy, which led us to develop large number of peptides of medical importance. Although, peptides are generally considered to be poor drug candidates because of their low oral bioavailability and propensity to be rapidly metabolized, yet development of synthetic single and/or linear peptides was found to be a remarkable advancement for drug discovery and therapeutics. In the past, a large number of potent peptide drugs have been developed as chemically defined or recombinant peptides. Although peptides have several advantages over small molecules and antibodies, there are some limitations associated with peptides which limit their use as a mainstream drug source. Firstly, they exhibit low potency and short half life, owing to rapid renal clearance, and lack *in vivo* stability due to protease degradation. Secondly, peptides also exhibit limited access to the intracellular space and can contain potential immunogenic sequences (random peptides) in general. Therapeutic peptides traditionally have been derived from three sources: (i) natural or bioactive peptides produced by plants, animal or human (derived from naturally occurring peptide hormones or from fragments of larger proteins); (ii) peptides isolated from genetic or recombinant libraries and (iii) peptides discovered from chemical libraries.

With the advancement of peptide synthesis technologies, improved productivity and reduced metabolism of peptides, along with alternative routes of administration, several bioactive peptides have been developed which are found to be highly functional with many serving as potent agonists and antagonists against several receptors involved in disease progression. Therapeutic peptides contribute significantly to the treatment of diabetic, osteoporotic, oncologic, gastroenterologic, cardiovascular, immunosuppression, acromegaly, enuresis, antiviral, obesity, antibacterial and antifungal indications (Table 1). On the other hand, a large number of immuno-prophylactic and immuno-diagnostic peptides have also been developed for several infectious diseases like HIV, Malaria, plague, influenza, anthrax and autoimmune diseases like Type 1 Diabetes etc. Although, there is an immense development in the field of peptide chemistry yet it has been associated with many difficulties which are related to the stability of peptides in the system, structure and sometimes immunogenicity of long peptides, that we are not interested. To use peptide as a therapeutic peptide, its biological activity, pharmacokinetic profile and low immunogenicity are crucial parameters. Various chemical modifications have been employed to overcome these limitations of peptides. Many cyclic peptides, pseudo-peptides (modification of the peptide bond) and peptidomimetics (non-peptide molecules) preserving the biological properties of peptides have been developed to increase their stability and bioavailability. However, targeted delivery of therapeutic peptides is still a challenging task. Many therapeutic peptides and peptide analogues do not have ideal delivery and release kinetics and hence new delivery system need to be developed which can ideally suited pharmacokinetics, bioavailability and toxicity profiles as well. In addition, peptide delivery profile/efficacy has been developed taking various pharmacokinetic parameters like extended release, pulsatile delivery, increased bioavailability and renal clearance. Furthermore, various routes of delivery (transdermal, pulmonary, nasal and oral) of peptide have been developed recently.

ACTH and derivatives	Corticorelin ovine triflutate (Ferring Pharmaceuticals Inc.), Corticorelin acetate (Celtic Pharma), Cosyntropin (Amphastar Pharmaceuticals, Inc), Seractide acetate (Clearsynth Labs Pvt. Ltd)
ACE inhibitors	ACE inhibitors Alacepril, Captopril (Dainippon Sumitomo Pharma Co.,Ltd.), Benazepril (Novartis), Cilazapril (AFT Pharmaceuticals Ltd), Delapril (Taizhou Huading Chemical Co., Ltd), Fosinopril (Camber Pharmceuticals Inc.), Imidapril (Chembest Research Laboratories Limited.), Moexipril (Schwarz'Pharma), Perindopril (Sandoz Limited), Quinapril (Shanghai Boyle Chemical Co., Ltd.), Ramipril (Sandoz/Novartis),
Antidiabetic agents	Exenatide (Amylin Pharmaceuticals), Liraglutide (Novo Nordisk), Pramlintide acetate (Arden pharmaceutical &chemical Co), GLP-1, Glucagon (Eli Lilly),
Anti-HIV compounds	Amprenavir (Chembest Research Laboratories Limited.), Atazanavir sulfate (Bristol Myers), Darunavir ethanolate (Tibotec), Fosamprenavir calcium (ViiV Healthcare), Indinavir sulfate (Merck),
Calcitonins	Salmon Calcitonin (novartis), Elcatonin Acetate (Shanghai Gene Biochemistry Co. Ltd.)
Cardiovascular	Bivalirudin Trifluoroacetate Hydrate (The Medicines Company), Eptifibatide (Millennium Pharmaceuticals), Rotigaptide (Zealand Pharma)
Cholecystokinin analogues	Ceruletide diethylamine (Pharmacia And Upjohn Co), Sincalide (Bracco Diagnostics Inc)
CNS	Cilengitide (Merck-Serono), Taltirelin hydrate (Jinan Haohua Industry Co., Ltd.), Ziconotide acetate (Zhengzhou Linnuo Pharmaceutical Co., Ltd.)
GHRH and analogue	Sermorelin acetate (GL Biochem), Somatorelin acetate (Ferring Pharmaceuticals Ltd)
GnRH and analogues	Buserelin acetate (Ningbo Hi-Tech Biochemicals Co., Ltd.), Gonadorelin acetate (GL Biochem), Goserelin acetate (AstraZeneca), Histrelin acetate (Xiangfan Goto Chemical Co.,Ltd.)
GnRH antagonists	Abarelix acetate (Chengdu Kaijie Biopharm Co. Ltd.), Cetrorelix acetate (Sigma), Degarelix acetate (Manus Aktteva Biopharma LLP)
Oxytocin, antagonist and analogues	Atosiban acetate (ProSpec), Carbetocin acetate (GL Biochem), Oxytocin (Tocris Bioscience)
Secretin	Secretin (ChiRhoStim)
Somatostatin (GHIH or SRIF) and analogues (agonists)	Depreotide trifluoroacetate (Creative Peptides), Edotreotide (Novartis PharmaAG), Lanreotide acetate (Ipsen Biopharmaceuticals, Inc.), Octreotide acetate (Bedford Laboratories), Pentetreotide (Novartis Pharma AG), Somatostatin acetate (GL Biochem)
Vasopressin analogues	Argipressin (Amdipharm Mercury Company Limited), Desmopressin acetate (Sanofi-aventis U.S. LLC), Lypressin (Shanghai Boyle Chemical Co., Ltd), Phenypressin (Creative peptides)
Antibodies	Abciximab (Centocor B.V.), Apcitide Tc-99m (Dr. Rentschler Biotechnologie GmbH), Bevacizumab (Genentech/ Roche), Catumaxomab (Fresenius Medical Care (UK) Ltd.), Crotalidae polyvalent immune Fab (Protherics Wales), Omalizumab (Genentech and Novartis), Ranibizumab (Genentech).
Anticancer	Bortezomib (Millennium Pharmaceuticals), Cilengitide (Pfizer), Histrelin (Arden pharmaceutical & chemical Co., Ltd), Goserelin (AstraZeneca), Stimuvax (Merck)
Immunotherapy	Cyclosporin (Novartis Pharmaceuticals Ltd), MPB8298 (Elan Corporation, plc).
Antibacterial	Daptamycin (Eli Lilly), Bacitracin (Novachem Ltd.), Colistin (Xellia Pharmaceuticals), Pexiganan (Genaera Corporation), Omiganan (Cutanea Life sciences)
Antifungal	Caspofungin (Merck), Micafungin (Astellas Pharma), Anidulafungin (Vicuron pharmaceuticals inc), Histatin (Sigma), Lactoferrin (Sigma).

Table 1: Some important peptides and peptide analogues.

Discovery of Novel Peptides

Peptides have emerged as one of the major classes of therapeutic molecules which have been developed as pharmaceutical drug entities by pharmaceutical and biotech companies in their search for drug discovery targets. Oxytocin was the first synthetic peptide discovered as therapeutic agent in 1953. The production of peptide therapeutic by recombinant means was introduced in 1974; however, the first peptide therapeutic to be approved for production using this method was human insulin in 1982. In recent decades, prevalent acceptance of peptide-based therapeutics by clinicians as well as patients and the advancement in peptide-related technologies has instigated researchers and pharmaceutical industries to develop all aspects of therapeutic peptides, including its discovery, safety, toxicology, clinical studies, manufacturing, formulation, delivery and marketing approval. Peptide-based therapeutic agents generally consist of natural peptides, peptide analogs and next generation peptides. The next generation of peptide-based bioactive compounds is presently on its way from fundamental research to clinical approval stage and eventually to the drug market.

To promote research and development and to help encouraging additional growth of peptide therapeutics, screening and discovery of novel and robust peptides is a very important step. Discovery of peptides as a therapeutic is a complex, time-consuming and multifaceted process and till now several strategies have been developed to screen and generate peptide-based pharmacologically active compounds. Conventional method of drug discovery involved synthesis of compounds in lengthy multi-step experimental studies followed by their *in vitro* and *in vivo* biological screening. The selected potential candidates were then investigated for their pharmacokinetic qualities for instance, stability, selectivity, toxic side effects, potential immunogenicity, and metabolism. Now, the advances in genomics, proteomics, and bioinformatics have revolutionized the process of drug discovery process by introduction of relatively efficient approaches like combinatorial chemistry, high throughput screening (HTS), virtual screening, de novo design and structure-based drug design. Combinatorial peptide library screening is the most widely used method in the drug discovery process. One such example is the phage libraries, which show potential leads in many therapeutic applications. Interestingly, in several cases, peptides screened by using this approach have been found to be more efficient, structured, and selective to targets but pharmacological description is still at their naive stage. Although, the traditional combinatorial approach can generate extensive small-molecule libraries (~105 compounds) [1,2] by virtue of unlimited functional group diversity, the major disadvantages associated with it is the time-consuming identification of active compounds from the pool and most importantly difficulty in decoding the synthetic libraries [3]. Later, more effective ways were developed in which the small molecule libraries were synthesized in the biological systems. This approach resulted into relatively more extensive libraries along with a straightforward and better detection of the active compounds [4-6]. However, the compounds synthesized by using this approach were having one serious drawback of making smaller and linear peptides that are often prone to degradation by the proteases [7,8]. To enhance the speed and effortless identification of a target-specific compound, more elegant approach was undertaken by integrating intracellular production of libraries of cyclic peptides, involving split-intein circular ligation of peptides and proteins (SICLOPPS), with *in vivo* screening system by different group of investigators [9-11]. In parallel, non-library based approach was also used in order to develop target interfering peptide sequence from predicted structure.

Detailed analysis of the peptide therapeutics over the past 15 years shows a remarkable trend toward clinical study of more peptide therapeutics especially non-canonical peptides. These are the peptides which have been modified to increase the shelf-life and their

ability to selectively target cells or penetrate the blood-brain barrier. Albiglutide (GlaxoSmithKline), dulaglutide (Eli Lilly & Company), NGR-hTNF (Mol Med S.P.A), and davunetide (Allon Therapeutics) are prime examples of such peptides that have progressed to Phase 3 studies. Recent report suggests that there are almost 80 peptides in the market, about 200 are in clinical phases and 400 are in advanced preclinical development stages [12,13], with over 75 per cent of these arriving in the last three decades. Sales for marketed peptide (glatiramer acetate (Copaxone®), leuprolide (Lupron®), octreotide acetate (Sandostatin®), goserelin acetate (Zoladex®), teriparatide (Forteo®), exenatide (Byetta®) therapeutics have also grown substantially in last decades. Towards the end of the previous century, the use of peptide drugs was very limited for the treatment of metabolic diseases, but recent reviews showing metabolism as the largest therapeutic field (25%) for peptides, followed by cancer (16%) and other clinical fields [14] revealing a significant and promising change in peptide growth trends.

Advantages and Disadvantages

The most important advantage of peptide therapeutics over all small molecule drug candidates is the structural relationships between the constructed peptide and the physiologically active parent molecules from which they are derived, an attribute that help in depicting the risk of unforeseen side-reactions. As compared to the small molecules that show the advantages of small size, low price, oral availability, easy synthesis, membrane-penetrating ability and stability, peptides are at a drawback [15-19]. However, in comparison to large molecules such as proteins and antibodies, peptides are still small. It is owing to its smaller size that the peptides can be easily synthesized, optimized, and evaluated to prevent any possible side effects. Moreover, the new drugs targeting protein–protein interactions often require larger interaction sites than small molecules can offer. The main benefit of peptides as therapeutics lies in their high activity, high specificity and affinity which are often in the nanomolar range, minimal drug-drug interactions, biological and chemical diversity etc. One of the important aspects of using peptides as drug candidate is their ability not to get accumulated in specific organs such as the kidney and liver, and thus help in minimizing their toxic side effects. In contrast, small molecules are not particular and can build up in various organs, ultimately leading to severe toxic side effects. Generally, small molecule peptides are not directly linked to severe side effects since they are composed of naturally occurring amino acids and which are metabolically endurable too. It is their dosage or delivery forms which are to be considered while evaluating the side effects.

Peptide drugs have also disadvantages mainly related to their *in vivo* instability such as, short half-life and low bioavailability, susceptibility to proteases, formulation and manufacturing challenges etc, resulting in restricted use until last few years. Major disadvantage of using peptides is the high production cost and the market price as compared to that of the small molecules. However, the cost of manufacturing peptides has been minimized with the rise in its production scale and efficiency due to developments in synthesizer, synthesis and purification strategies.

Chemical Synthesis of Peptides:

Concept of synthetic peptides was started in the first halves of twentieth century. After 1950, there had become a significant development in the field of peptide chemistry in order to define wide role of peptides in all life processes. It had been more interesting to analyse different peptides by the development of sensitive analytical techniques. It was the group of Friedrich Wessely (1897-1967) which systematically synthesized peptides using N-carboxyanhydride (NCA) method. The invention of solid phase peptide synthesis (SPPS) by Bruce Merrifield in 1963 was the revolutionary development in peptide synthesis. This led to the acceleration of peptide synthesis resulting in production of thousands of peptides and peptide analogues in a short time. At the time it was clear that peptides were a very important biologically active class of naturally occurring compounds. It involved the stepwise addition of protected amino acids to a growing peptide chain which was attached to a solid resin particle. This approach provided an easy procedure whereby reagents and by-products can be easily removed simply by filtration. The advantages of the new method were speed and simplicity of operation [20]. Prior to the development of SPSS, peptides were synthesized via solution phase method, which was tedious and required high level of safety and skill. SPPS also allows the synthesis of natural peptides which are difficult to express in bacteria, the incorporation of unnatural amino acids, peptide/protein backbone modification, and the synthesis of D-proteins, which consist of D-amino acids. In 1971, R. B. Merrifield synthesized and characterized Ribonuclease a, linear polypeptide of 124 amino acid residues, by the solid phase method using t-Boc (N-tert-butoxycarbonyl) protecting amino acids on styrene-divinyl-benzene resin. Basic concept of SPSS is to covalently attach the first amino acid to an insoluble support (Resin) and to extend the peptide chain from this support bound residue and desired length of peptide is completed through a series of coupling and deprotection steps. After completion of the desired length of peptide, the peptide is removed from solid support by using appropriate cleaving reaction. One of the most important advantages of SPSS is that the growing peptide chain can be washed to remove uncoupled amino acid by using suitable solvent, which makes sure that unwanted amino acids are not attaching in the growing peptide chain [21].

SPSS normally proceeds in the C→N direction. Solid support (resin) holds entire growing peptide chain. A linker, in between the resin and the first amino acid, can be attached to avoid stearic hindrance and facilitate smooth synthesis of peptide. All amino acids used are orthogonally protected with temporary protecting group at N-terminal and permanent protecting groups at side chain with C-terminal free to form peptide bond with N- terminal of previously attached amino acid, which make sure that peptide is growing at C- to N-terminal orientation. Once the first amino acid is attached to the resin, its N-terminal is deprotected using appropriate solvent and next amino acid is attached. Solid supports for SPSS generally increase significantly in size, i.e. "swell" when solvated and reactions take place throughout the support (interior as well as surface) [22,23]. Polystyrene, cross-linked with divinylbenzene has been widely used [24]. Polyethylene glycol (PEG) grafted onto polystyrene or ethylene oxide polymerized onto the polystyrene (Tentagel and ArgoGel) [25,26]. Several other polymeric supports are now available which can be derivatized with functional groups to produce a highly stable linkage to the peptide being synthesized [27] and peptides with different functionalities in the terminal carboxyl group (i.e. amide, acid, thioester). Some more resins can be used such as p-methoxybenzhydrylamine (MBHA), 4-hydroxymethylphenylacetamidomethyl (PAM), 4-(2', 4'-dimethoxyphenyl-amino methyl) -phenoxymethylpolystyrene (Rink), 2-chlorotrityl chloride, 4-alkoxybenzyl alcohol (WANG) and diphenyldiazomethane based resins.

In the last few decades, several protecting groups have been proposed to make peptides synthesis easy and to avoid use of harmful chemicals in the synthesis [28]. Currently, two main schemes of protection, which are known as t-Boc/Bzl and Fmoc/tBu have been used [29]. The Boc-benzyl strategy depends on graduated acid lability. The N-terminal Boc group is removed by Triflouroacetic acid (TFA) in dichloromethane (DCM). TFA treatment produces isobutylene, carbon dioxide, and a protonated amine, which must be neutralized before the next coupling. In case of t-Boc strategy, side chain protecting groups need to be TFA resistant in order to prevent premature cleavage of side chain. Removal of side chain protecting groups occurs with strong acids such as hydrofluoric acid (HF) or

trifluoromethanesulfonic acid (TFMSA), in presence of suitable scavengers, anisole [30,31]. The Boc scheme has proven to be successful for peptide synthesis for decades, but the use of dangerous chemicals, harsh final acidic cleavage conditions and specially designed work place with skilled handler, provided an impetus to researchers for the development of milder alternatives (Figure 1).

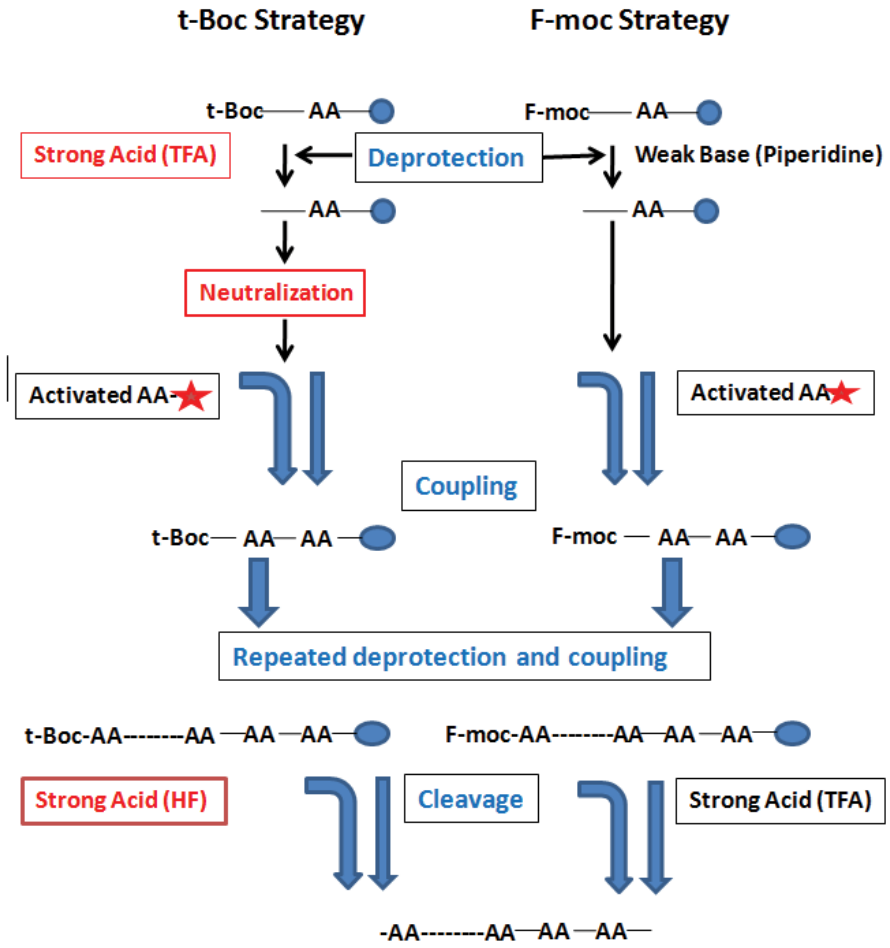


Figure 1: Schematic diagram showing Synthesis of peptide using t-Boc and Fmoc chemistry.

In Fmoc/tBu, the Fmoc (9- fluorenyl methoxycarbonyl) group is used for the protection of the N-α amino group and the tert-butyl group for the side chains of several amino acids [32]. In Fmoc strategy, N-terminal Fmoc group can be removed by a variety of secondary amines, usually 20-50% piperidine in *N*, *N*-dimethylformamide (DMF) and there is no neutralization needed after deprotection while tert-butyl groups which are used to protect side chains of Fmoc amino acids, are cleaved by trifluoroacetic acid which also cleaves the peptide from the resin. Moreover, use of DMF as a solvent improved the yield of peptide. The need of milder conditions and easy handling of chemical used in Fmoc strategy makes it more popular than t-Boc strategy [Figure 1].

Moreover, synthetic peptides also showed immunological response in animal models. But, being small in size and simple structure, they are not found to be good immunogen until incorporated with some carrier/adjuvants. Tam et al in 1988 developed a method of synthesizing multiple antigen peptides (MAP) in an oligomeric form and as a macromolecule by the solid-phase peptide synthesis method. The MAPs consist of multiple copies of peptides circumsporozoite proteins of two species of malaria that are synthesized as single units on a branching lysyl matrix using a solid-phase peptide found to be more immunogenic as compared to a monomeric peptide [33-35]. In 1990, Tam et al. synthesized MAP containing four copies of peptide antigen [B(4), T(4), BT(4) and TB(4) epitopic arrangement] separately on a tertbutoxycarbonyl (Boc)-Ala-Pam resin using three Boc-Lys(Boc) amino acid as a starting point. Similarly, they also synthesized MAP models containing eight copies of the antigens [T(8), B-(8), BT-(8), and TB-(8)], the core was synthesized with seven lysines prepared from branching with three levels of Boc-Lys(Boc). In the MAP models containing eight copies of B or T antigens and one copy of T or B antigen [B-(8)-T, T(8)-B], the B or T antigen was first synthesized linearly on the Boc-Ala-Pam resin and then branched with three levels of lysine to give subsequent eight copies of T or B antigens. All peptides and MAPs were cleaved from the resin supports by low/high HF procedure (Figure 2) [36].

In 1992, Jean-Paul Briand et al. developed several MAP systems containing eight copies of 6-15 residue-long peptides derived from selected regions of various proteins. They observed that immunogenicity of MAPs was found to be superior than the same peptides linked to carrier protein by means of conventional conjugation procedures [37]. Multiple antigen peptide (MAP) consisted of a malaria T-cell epitope or of a human universal tetanus toxin helper T-cell epitope collinearly synthesized with B-cell epitopes from the circumsporozoite proteins of different malaria parasites stimulated specific T-cell clones [38]. Also, MAP system, consisted of an oligomeric branching lysine core to which B and T cell epitopes of meningococcal class 1 protein were attached as octameric and tetrameric dendron like structure, was synthesized and confers some conformational stability and predominance of antibodies directed towards conformational epitopes [39]. In the year 2000, Robert A. Boykins et al. developed a MAP containing multiple epitopes of human malaria parasite and the Tat protein of HIV type-1 (HIV-1-Tat) using Fmoc strategy [40]. Tetra-branched MAP containing four copies of peptide of the

24 amino acid extracellular domain of matrix protein 2 (M2e-MAP) of H5N1 virus strain VN/1194, a remarkably conserved across influenza A subtypes, was synthesized on $[Fmoc\text{-}Lys(Fmoc)]_2$-Lys-Cys(Acm)-βAla-Wang Resin using Fmoc chemistry and cleaved of the peptide from the resin by treatment with trifluoroacetic acid (TFA), DTT, water, and triisopropylsila (TIPS) in the ratio 88:5:5:2, respectively. Crude peptide was purified by reversed phase high-performance liquid chromatography (RP-HPLC) and characterized by amino acid analysis and matrix assisted laser desorption ionization mass spectrometry (MALDI-MS). M2e-MAP vaccine induced strong MAP specific antibody response in murine model [41]. Fujiki T et al. in 2010, designed octavalent MAP containing eight copies of epitope of human tumor necrosis factor-α (TNF-α) synthesized on seven residues containing lysine backbone and *in vitro* developed TNF-α specific monoclonal antibodies in human peripheral blood mononuclear cells (PBMCs) [42].

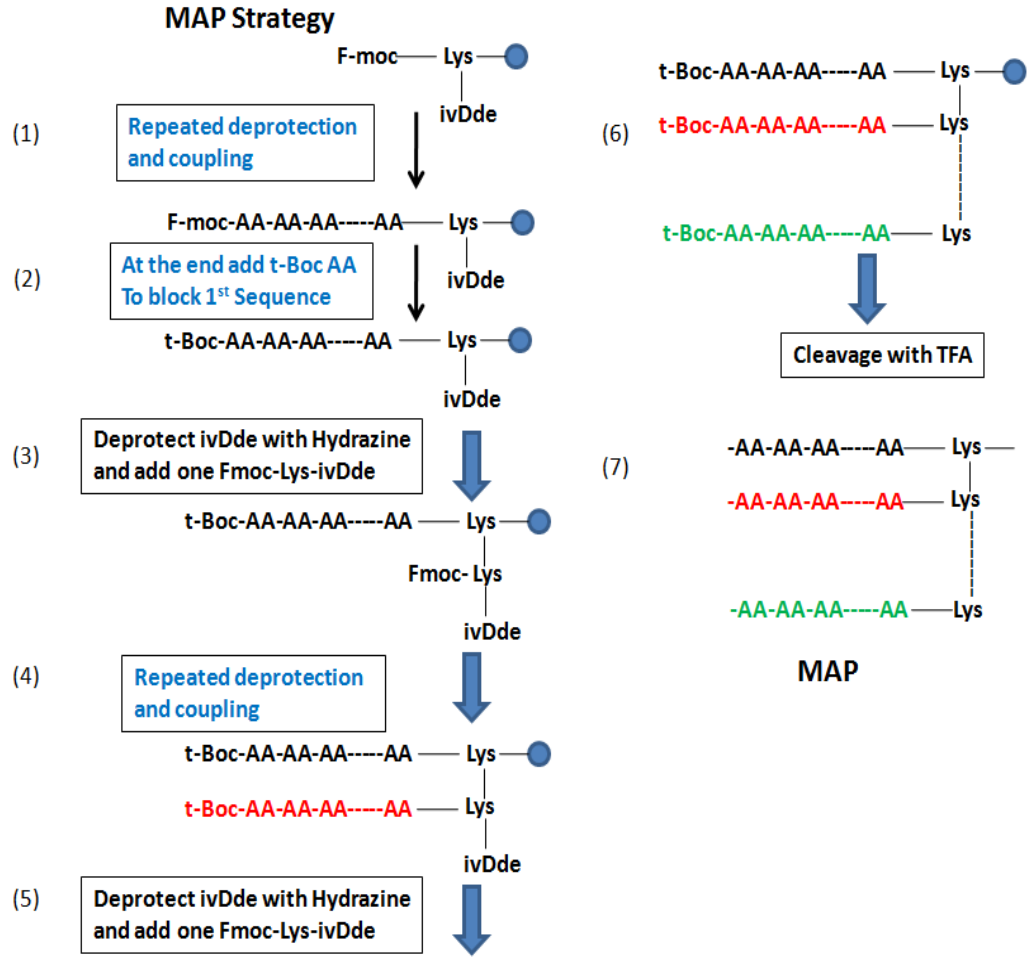


Figure 2: Schematic diagram showing Synthesis of Multiple Antigen Peptide (MAP) using chemistry.

Two chimeric multiple antigenic peptides (CMAP5 and CMAP8) and 4 monomeric linear peptides (LP1M, LP1N, LP1O, LP2) derived from the transmembrane proteins (gp41) of HIV-1 and HIV-2 were developed on a lysine core. CMAP5 and CMAP8 seropositivity with HIV-1 and 2 thereby, showing highly sensitive and specific assays for the detection of infections caused by HIV-1, including group M, N, and O, and HIV-2 [43]. Recently, our laboratory developed a MAP containing three B, one T cell epitopes of F1-Ag of *Yersinia pestis* on Fmoc–Gly–HMP–Tantagel resin using Fmoc chemistry. Fmoc-Lys (ivDde)-OH was incorporated at the beginning of the synthesis for each branch to serve as the branching point. Lysine is a unique amino acid having two amino terminals and thus it can support two N-terminal coupling reactions, thereby providing two groups with minimum stearic hindrance. ivDde is a protecting group which is resistant to piperidine, which is used to deprotect Fmoc group during synthesis. First sequence was added at the N-terminal end of first lysine and after completion, blocked with t-Boc protected amino acid to prevent further chain elongation. Use of t-Boc amino acid, being resistant to piperidine reaction, at the end of each sequence made possible to add new sequence without extending existing one. After completion of one sequence, ivDde group on ε-amino group of first lysine was deprotected by using hydrazine and another lysine (ivDde) was added at the same point. Next sequence was added on a-amino group of second lysine and subsequently completed the desired sequence before final deprotection with TFA [Figure 2] [44].

Biotechnology Approach for Peptide Production

Peptides are of growing interest not only as therapeutics but also as transporters of non-peptidic molecules such as small molecule active pharmaceutical ingredient (API), cytotoxic or imaging agents. Increasing demand of peptide as therapeutics and major challenges associated with it has not only led to abundant innovations in stabilization or formulation technologies but also in manufacturing techniques. Producing therapeutic peptides is a complex and time-consuming process and can take several years just to identify it, determine its gene sequence, and validate a biotechnology process to manufacture it. There are certain challenges especially associated with the large-scale synthesis of peptides such as, to meet the production demand, proper strategy of synthesis applicable on all scale, downstream processing and isolation, quality control, regulatory implications, and production economics, which are needed to be

considered for their persisted existence in drug markets. These challenges are being actively tackled by intense research and development and working on alternative technologies.

Several technologies for the production of peptides are now available which includes isolation from natural sources [45], production by recombinant DNA technology [46], production in cell-free expression systems [47], production in transgenic animals [48] and plants [49], production by chemical synthesis [20,50] and production by enzyme technology [51]. The type of manufacturing technologies to be opted is primarily determined by the size of the peptide molecules. Before the advances in synthesis technologies, peptides were extracted from natural sources for instance; human growth hormone (hGH) and insulin were collected from human corpses and slaughtered pigs, respectively. The resultant peptides were found to be expensive, less in availability and restricted to native peptides so, alternative methods were developed that can be applied to manufacture native as well as synthetic peptides. Even though chemical synthesis is the most developed technology used in peptide synthesis, multi-step synthesis process, high synthesis cost, increase in length and complexity, increasing problems with chain aggregation, low yield, high toxic effluents, and lack of specificity are severe drawbacks that can in principle be successfully overcame by using biotechnological tools and techniques. In recent years, peptide synthesis using recombinant DNA technology and biocatalysts has emerged as a new hope in the field of peptide drugs and their implementation can greatly affect the cost and scalability of pharmaceutical peptides.

Recombinant DNA technology offer several advantages at large production scales and is specifically suitable for the synthesis of large peptides and other hormones [52]. Major steps involved in recombinant production of pharmaceutical peptides includes selection of efficient expression systems, such as *Escherichia coli* [53], *S. aureus* [54], insect cells [55], transgenic mammals [56] and transgenic plants [57], for over expression to get larger and purer qualities of the peptides followed by fermentation, purification of fused protein, cleavage of fused protein using endopeptidases, purification of peptide, capping of peptides (N-terminal acetylation which is *in vivo* catalyzed by N-acetyltransferases or C-terminal amidation by peptidyl-glycine alpha-amidating monooxygenase (PAM)), purification of capped peptides and formulation of final product. Although, different recombinant technologies have been developed for the production of peptides, however, inefficient recombinant expression of peptides owing to its susceptibility to degradation in the cytoplasm mainly due to their relatively smaller size and lack of tertiary structure, their toxicity against the bacterial host cells and post-translational modifications particularly in case of synthesis of peptide hormones, remains a significant challenge. Direct-expression technology along with a technology for *in vivo* amidation using PAM has recently been developed for the efficient and cost-effective production of peptide hormones. To overcome the problem of degradation, peptides can be expressed in fusion with a larger protein which in turn helps the resultant protein from proteolysis and in accumulating in cytoplasm in soluble or insoluble forms owing to their larger size [58]. However, difficulty in liberation of the fusion counterpart from the peptide by chemical or enzymatic cleavage, its purification and overall low product yield are the major challenges associated with it that decreases the cost-effectiveness of the system. Thus, for wide application of recombinant DNA technology, an extended and costly research and development is required as it often remains unachievable mainly due to the less efficient expression systems and difficulties in product extraction, purification and recovery [59].

In enzymatic peptide synthesis, proteases (endo- and exo-) are the most prevalent class of enzymes that are used as biocatalyst and are chosen based on their specificity against amino acid residues. Furthermore, it is their ability to biocatalyze reactions in non-aqueous media such as organic solvents [60,61], supercritical fluids [62,63], eutectic mixtures [64], solid-state [65,66] and, more recently, ionic liquids [67-70] that has expanded their applications in peptide synthesis [71,72]. In general, there are two mechanisms used in enzyme-catalyzed peptide synthesis and are based on type of the carboxyl component used, one is thermodynamically controlled synthesis (TCS) and another is kinetically controlled synthesis (KCS) [73]. In the TCS approach, this component has a free carboxyl terminus, and the peptide bond formation occurs under thermodynamic control. TCS is the reverse of the hydrolytic cleavage of peptide bond in presence of proteases with the formation of a acyl intermediate, where, catalyst do not alter the equilibrium of the reaction and simply enhances the rate of the reaction [74]. Since, the formation of acyl intermediate is the rate limiting step in TCS, the use of an acyl donor with free carboxylic group and any type of proteases are the major considerations. The main disadvantages of this process are high requirements of biocatalysts, need of highly précised reaction conditions to displace the equilibrium towards peptide bond formation, low reaction rates and low product yield [75]. Displacement of equilibrium can be achieved by modifying the pH and medium composition [76,77], however maintaining the properties of the biocatalyst, substrate and products is challenging. In the KCS approach, the carboxyl component is used in an activated form, mainly as an ester derivative, and the synthesis occurs under kinetic control [73]. Unlike TCS, serine or cysteine proteases (papain, thermolysin, trypsin and chymotrypsin) are used in KCS, because the main function of enzyme is to act as a transferase destined to catalyze the transfer of an acyl group from the acyl donor to the amino acid nucleophile through the formation of an acyl-enzyme intermediate. Enzymatically synthesized peptides not only have application in pharmaceuticals but have also been explored in agrochemicals, human and animal nutrition [78-82]. This technology has also some limitations as it is specifically suitable for synthesis of very small peptides [60] and cannot be used for the production of peptides more than 10 residues in length. However, enzyme stereospecificity, mild reaction conditions, minimum side chain protection and avoidance of racemizations can overcome some negative aspects of chemical method of peptide synthesis.

Peptide Purification and Isolation Of Purified Peptides

Before being formulated into pharmaceutically active products, the post-synthesis mixture undergoes purification and isolation process to obtain the desired peptide product. This is most crucial step that must be carefully assessed as it plays a vital role in determining the optimal economy of manufacturing process and eventually the utilization efficiency of peptides as therapeutic. It is a complex process which involves several steps to ensure that the desired products meet the quality requirements set for the compound to be purified with negligible impurities and are achieved cost-effectively. In general, the type of techniques to be used for purification purpose relies on the properties of the peptide and the types of impurity present. Today, most commonly used standard technique for isolation is preparative high performance liquid chromatography (HPLC) [83,84] whereas, ion-exchange chromatography, gel permeation chromatography and medium or high pressure reversed phase chromatography are used for purification of peptides [85,86] but other methods like counter current distribution [87] and partition chromatography [88] were used in the past. In recent years, ultra performance liquid chromatography (UPLC) has become a technique of choice for the separation of various pharmaceutical related small organic molecules, proteins, and peptides. Using UPLC, it is now possible to run separation process using shorter columns, and/or higher flow rates for increased speed, with superior resolution and sensitivity [89-91].

In ion exchange chromatography (IEC), separation is dependent on the ionic interaction between the support surface and charged groups of the peptide (cations and anions). High flow-rates, efficiency and high mechanical strength of the ion exchange material are the most desirable factors for large-scale purification process. Low mechanical strength may have a negative impact on recovery and

separation efficiency. However, use of an organic modifier in the mobile phase can solve this problem. In IEC, eluent is often a volatile salt, which can easily be removed by lyophilization, reverse osmosis or solid phase extraction. The Gel permeation chromatography (GPC) is highly efficient technique for separation of polymeric forms of peptides and for desalting of peptide solutions and separates molecules on the basis of size exclusion. Major concerns associated with gel permeation chromatography are their low capacity and the relatively low flow-rates. Acetic acid is the most commonly used eluent for industrial scale purification of peptides by GPC. Reversed phase chromatography (RPC) has been found to be the most powerful method for peptide purification which utilizes hydrophobic interactions as the main separation principle. For large-scale purification, shape and size of the particles of the stationary phase are important considerations as it directly influence column efficiency. In small-scale RPC, organic modifiers such as acetonitrile, methanol, isopropanol etc are used as an eluent, however these are not as efficient while considering for industrial scale. Therefore, in addition to separation efficiency, other important aspects such as impact of modifier on environment and process economy should also be considered. Substituting these modifiers with ethanol as an eluent has been proven to be the most efficient way to overcome this situation.

In a typical purification process, the crude mixture is first subjected to a "capturing" step which can be initially achieved by application of ion exchange chromatography. For obtaining purity of higher degree, "polishing" step is performed by applying the peptide solution to a column packed with reversed phase resin. Thus, combination of two or more complementary methods can lead to robust purification processes. Mass spectrometry data and amino acid sequence analysis are acquired to validate the identity of the target peptide. Prior to isolation of the purified peptides, peptide solution needs to be concentrated, which is normally carried out under reduced pressure. This step assists not only in obtaining the desired peptide in the solid concentrated form but also in controlling the quality traits such as content of water, counter ion and residual solvents. In case of large-scale purification process, where high volumes of eluent is challenging, reverse osmosis technique has emerged as the best alternative to reduced pressure method with a dual advantage of concentrating the peptide solutions as well as removing the low molecular weight salts and organic solvents. For isolation of the purified peptide, lyophilization or freeze-drying is the most commonly used method. However, hydrophobic peptides present significant purification challenges as they are not readily soluble in typical purification buffers. Currently, promising alternatives to freeze-drying, such as large vessel precipitation and spray drying, are under consideration, but spray drying can further create problems for the peptides that are not thermostable. With reference to the purification and isolation process, in-depth knowledge of peptide production methods, identification of critical steps and parameters of the purification and predetermined limits of critical process parameters are essential considerations for optimized product yield and purity in the way to enable cost-effective and faster commercialization of peptide-based therapeutics.

Targeted Delivery and Pharmacokinetics of Therapeutic Peptides

Delivery of peptides and therapeutic profile are two of the important parameters for peptide drugs which can be improved by generating peptidomimetics with improved stability and permeability. In addition, targeted delivery of drug had been improved in last decades. Targeted drug delivery has been influenced by a wide range of structure-activity relationships, peptide analog generation to impart stability in the system against proteases and increased bioavailability, and novel formulations to target optimal therapeutic dosing requirements. With the development of a wide range of biodegradable polymers and nanoparticles based delivery vehicle, it has become possible to administer peptide drug efficiently and to monitor drug release, bioavailability and toxicity of peptide drug. Several delivery vehicles in a large number of diseases have been studied recently. Jin et al. 2011 showed that the oral administration of Trimethyl chitosan chloride (TMC) nanoparticles based delivery of CSKSSDYQC (CSK) peptide showed sufficient effectiveness as goblet cell-targeting nanocarriers for oral delivery of insulin and subsequently produced a better hypoglycaemic effect in diabetic rats [92]. GRN1005 is a novel peptide-drug conjugate composed of paclitaxel covalently linked to a peptide, angiopep-2, that targets the low-density lipoprotein receptor-related protein 1, showed high dose tolerance in heavily pre-treated patients with advanced solid tumors, including those who had brain metastases and/or failed prior taxane therapy [93]. Moreover, an injectable, phase sensitive, in situ forming, implantable, poly (D, L-lactide-co-glycolide) (PLGA), a biodegradable polymer based delivery system was developed for enfuvirtide, a therapeutic peptide used in the treatment of HIV infection. This PLGA based delivery system maintained required drug plasma concentration and was found to be biocompatible with the animal tissue [94]. An integrin α(5)β(1) antagonist, N-acetyl-proline-histidine-serine-cysteine-asparagine-amide (Ac-PHSCN-NH(2)) peptide was used as a novel homing peptide to deliver doxorubicin (PHSCNK-PL-DOX) loaded liposomes and showed stronger tumor inhibition and prolonged survival and less toxicity of mice bearing B16F10 tumors [95].

PEGylation is defined as the covalent attachment of poly(ethylene glycol) (PEG) chains to bioactive substances. It has been extensively studied recently, and the number of agents newly developed with PEGylation is increasing continuously. Dapp et al. 2011, synthesized PEGylated bombesin (BN) analogs for imaging of tumors that overexpress gastrin-releasing peptide receptors (GRPR) and showed increased stability of the analogs, improved their pharmacokinetics, and enhanced the tumor retention [96]. A series of biodegradable polydepsipeptides based new triblock copolymers, poly (ethylene glycol)-poly(L-lactide)-poly(3(S)-methyl-morpholine-2,5-dione) (mPEG-PLLA-PMMD) have shown lower CMC value, positive-shifted zeta potential, better drug loading efficiency and stability as compared to the mPEG(2000)-PLLA(2000) diblock copolymers for paclitaxel (PTX) delivery [97]. Sustained and targeted delivery of paclitaxel to tumor sites, truncated fibroblast growth factor fragment-conjugated PEGylated liposome showed promising potential as a long-circulating and tumor-targeting carrier system [98]. Cationic Aminomethylene peptide nucleic acid (am-PNAs) having pendant aminomethylene groups at α(R/S) or γ(S) sites on PNA backbone have shown to stabilize duplexes with complementary cDNA and demonstrated to effectively traverse the cell membrane, localize in the nucleus of HeLa cells, and exhibit low toxicity to cells [99]. LTVSPWY peptide-modified PEGylated chitosan (LTVSPWY-PEG-CS) modified magnetic nanoparticles selectively taken up by SKOV-3 cells overexpressing HER2 when cocultured with HER2-negative A549 cells was found to be a promising agent for early detection of tumors over expressing HER2 and further diagnostic applications [100]. Curcumin, the principal curcuminoid of the popular Indian spice turmeric, has a wide spectrum of pharmaceutical properties such as antitumor, antioxidant, antiamyloid, and anti-inflammatory activity. Curcumin was entrapped in methionine-dehydrophenylalanine, a novel self-assembled dipeptide NPs, and showed enhanced toxicity towards different cancerous cell lines and enhanced curcumin's efficacy towards inhibiting tumor growth in Balb/c mice bearing a B6F10 melanoma tumor [101]. Low molecular weight polylactide (LMW PLA) and V6K2 peptide showed tumor cell invasion with Doxorubicin (DOX) displayed higher tumor toxicity than PLA-EG NPs and lower host toxicity than the free DOX in C3HeB/FeJ mice inoculated with MTCL syngeneic breast cancer cells [102].

Antimicrobial peptides have shown to be attractive therapeutic agents with unique mechanisms of action, to be applied against several bacteria. However, the uses of antimicrobial peptides as therapeutics attract great concern due to their high hemolytic activity which cancels out the safety requirements for a human antibiotic. Therefore, efforts have been made to develop new class of potent antimicrobial peptides with minimal or no toxicity over erythrocytes. Several antimicrobial peptides (AMPs) have been studied and

used for therapeutic purposes. Recently, peptides (melectin, lasioglossins, halictines and macropin) and their analogs were isolated from the venom reservoirs of wild bees and characterized. These peptides showed high antimicrobial activity against Gram-positive and -negative bacteria, antifungal activity and low or moderate hemolytic activity. These peptides showed various degree of cell toxicity with normal and cancerous cell line [103]. Gomesin, a cationic antimicrobial peptide purified from haemocytes of the spider *Acanthoscurria gomesiana*, treatment effectively reduced *Candida albicans* in the kidneys, spleen, liver and vagina of infected mice [104].

Cell penetrating peptides, generally categorized as amphipathic or cationic peptides, are of increasing attention as a non-invasive delivery technology for macromolecules. Delivery of large numbers of diverse nature of therapeutic peptides has been attempted using different types of cell penetrating peptides (CPPs) *in vitro* and in vivo. The SynB peptides (RGGRLSYSRRRFSTSTGR), a family of cell-penetrating peptides showed charge-mediated blood–brain barrier selectivity [105] and has been used extensively in cationic cell-penetrating peptide vector-mediated strategies to deliver a large number of small molecules as well as proteins across cell membranes *in vitro* and across the blood–brain barrier *in vivo* [106,107]. Moreover, SynB-PEG-GS (Gelatin-siloxane) nanoparticles had shown to be a good biocompatibility with brain capillary endothelial cells and higher cellular uptake than that for PEG-GS nanoparticles [108]. CADY-1, an amphipathic peptide, is capable of forming complexes by self-assembly, and possesses cell-penetrating activity, and form a stable complex with doxorubicin. CADY-1 and doxorubicin complex extended the blood residence time of doxorubicin in a similar fashion to that of liposomal doxorubicin. In addition, the complex was capable of carrying doxorubicin across the cell membrane, which increased the therapeutic index of doxorubicin. Also, CADY-1/doxorubicin complex exhibited better tolerance and anti-tumour activity in experimental animals [109]. Recently, a hepatocarcinoma-binding peptide (A54 peptide) with PEGylated stearic acid grafted chitosan (A54-PEG-CS-SA) micelles directed doxorubicin into human hepatoma cells (BEL-7402) *in vitro*, and high distribution ability to liver and hepatoma tissue in vivo. A54-PEG-CS-SA micelles mediated targeted doxorubicin delivery to hepatoma cells suppressed tumor growth more effectively and reduced toxicity compared to commercial adriamycin injection [110]. Recently, EAK16-II, an ionic-complementary, self-assembling peptide, has been found to stabilize ellipticine in aqueous solution and enhance antitumor activity of ellipticine encapsulated in EAK16-II (EAK-EPT) was evaluated *in vitro* and *in vivo* without causing any side effects [111]. Cell-penetrating peptides (CPP) have been used to enhance cellular uptake of Morpholinos, a potential anti-bioterrorism agent for inhibiting replication of deadly Marburg viral infection [112].

Advances and Challenges in Peptide Based Therapeutic Approaches

Challenges

Despite increasing rate of approval and several advantages over small molecules, there are still some significant challenges associated with the peptides which limit its applicability as drug candidate and finally affect its commercial status in drug market. One of the major challenges is that most of the peptides cannot be administered orally due to the poor oral availability of the peptides, attributed mainly to its rapidly inactivation and poor assimilation through the intestinal mucosa. These characteristics of peptides substantially restricts the use of the most convenient and comfortable way of drug administration i.e. oral delivery, thus leaving option for alternative delivery routes such as subcutaneous or intravenous, nasal, sublingual, buccal, transdermal, pulmonary etc. as well as novel delivery systems like using liposomes [111-116]. However, there are a number of challenges when considering other routes of administration as different routes have been found to be linked to changes in the pharmacokinetic profile as well as the biological activity of the peptides [117]. The liposomal formulation technology frequently suffers from the active pharmaceutical ingredient (API) encapsulation efficiency into liposomes. Improving the stability and half-life of therapeutic peptides are other important issues as the gastric acid in the stomach and proteases (peptidases) in the intestinal, tissue and serum could easily degrade peptides. Likewise, peptide antigenicity can result in severe immune responses; therefore, there are also significant challenges in screening and development of novel peptides. Another challenge associated with manufacturing of peptide-based drug is the inability of regulatory authorities from different regions to agree on a common guideline defining the minimal level of peptide impurities present in peptide therapeutics. It directly affects the process designing meant for producing high quality products and reduces the economic feasibility by increasing the cost of the final products as compared to the traditional molecules. It is evident from these facts that there is still a long way to go to find novel and effective peptides with high commercial potential.

Advances

In order to face the peptide-associated challenges and to sort out the problems related to their commercialization, several advances have been made in the field of peptide based drugs to develop novel peptide therapeutics. The phage display is the recently developed technology which may pave an improved way for identification and discovery of potential peptides. Additionally, bioinformatics and systematic biological based strategies are increasingly being supported in boosting search for novel peptide drug candidates and are solely based on the available information and comprehensive data facts. Discovery of new generation peptide-based vaccine is another feather in the development of peptide therapeutics and may offer a wide product range in global market [118].

Poor bioavailability of peptides is one of the major concerns in development of an effective and robust therapeutics [119,120]. Attachment of polyethylene glycol (PEG) moieties to the peptides is a common strategy that have been used to improve the bioavailability of peptides with the additional benefits of improved stability from proteolytic degradation and protection from recognition by the immune system [121]. PEG-Intron and Pegasys, PEGylated interferon alpha-2b and alpha-2a, respectively, are examples of such peptides therapeutic currently being used in therapy of hepatitis C infections [122]. However, the use of an efficient targeting method (PEG and ligands (e.g. antibodies, mannose) attached to peptide–liposome formulation) or slow releasing drug delivery system (such as liposome encapsulation of peptides) has markedly reduced the requirement of such bulky modifications [115,116].

Stability is another major consideration in peptide therapeutics and several advances have been made in this field in order to modulate the *in vivo* stability. With the recent advances in technologies, the stability and selectivity of the peptides can be enhanced by modifying its structure and/or naturally occurring ligands to make it as long-acting analogs with high binding affinity and high receptor subtype-specific selectivity. For example, natural somatostatin has the ability to target all the five somatostatin receptor (SSTR) subtypes, but its synthetic analogs such as L797 &591, octreotide & lanreotide, and BIM23268 show preference for SSTR1, SSTR2, and SSTR5-subtypes, respectively [123]. To improve the half-life of peptides and to reduce the immunogenicity, their respective moieties have been successfully fused to human albumin (ex-albiglutide) or an engineered human IgG4 Fc heavy chain (ex-dulaglutide). Furthermore, via binding to the receptor FcRn, the albumin and Fc-containing molecules were made completely protected from degradation in cells. Beside using traditional methods of backbone modification (involving the modification of the N-terminus and C-terminus of

the peptides through acetylation and amidation, respectively, and/or the replacement of amino acids with unnatural amino acids or D-amino acids, or peptide bond modification using beta-aminoacids) [124-129], peptide resistance to proteases have been successfully engineered using recent technology such as Phage display by generating synthetic peptides with little sequence homology to the parent protein however, retaining the same biological activity [130], or other *in vitro* library selection system like CIS display [131-133]. For example, Hematide, an erythropoiesis stimulating agent, was developed based on this approach and was found to be non-immunogenic as well as stable in serum [134]. Various attempts have been made to improve intracellular stability of the peptide, based on constraining peptide structures [135,136]. For instance, attaching short dimerizing peptides to carboxyl and amino end of 18-mer peptides resulted into a stable conformation of monomeric tertiary structures [137,138]. Cyclic peptides have been rationally synthesized that have been found to possess not only *in vitro* and/or *in vivo* stability profiles but also high selectivity profile for bacterial rather than mammalian cell membranes [139-140]. Furthermore, incorporating peptides within the scaffold of a larger protein such as thioredoxin [141] has been employed as an alternative way to stabilize it. Some attempts have also been made to increase the clearance rate of peptide therapeutic agents by using glycosylation strategy. It can be achieved during chemical synthesis process by using glycosylated amino acids to produce glycopeptides or by conjugation of a carbohydrate unit to the full-length peptide [142].

Other approaches involving receptor-specific drug delivery carriers and nanoparticles as a carrier conjugated with peptides have been employed in order to improve the pharmacokinetics of the peptide based drugs and its assimilation in intestine [143-145]. This strategy could significantly help in overcoming the most difficult challenge associated with delivery of peptide drugs through oral route. Development of a new family of cell-penetrating peptides (CPPs) such as penetratin, M918, TP10 etc. exhibiting high efficacy and good internalization capacity has more or less solved the problem of poor oral bioavailability of the peptides by making it possible to deliver it through cell membrane with less damage [146-149]. New generation approaches in receptor-targeted therapeutics together with receptor-specific drug delivery carriers will further help in increasing the internalization of peptide based drugs, their selectivity to target cells, as well as their efficacy thus, reducing the toxic side effects and multi-drug resistance [150-153]. The introduction of smart linkers that demonstrate stability towards blood plasma but intracellular lability will lead to target-specific activity, which might successfully decrease side effects.

Recent progresses in manufacturing stream such as improved synthesizer, economical synthetic and purifying approaches and effective formulation methods, will certainly help in reducing the manufacturing cost by improving the scales of production [154]. However, the process development remains an important factor to be considered for costs and time efficient peptide drug development. With the gradual development in stabilization techniques aimed at improving the pharmacokinetic properties of peptides, and continuous advances in manufacturing, delivery and formulation technologies, peptide-based drugs has matured a lot over the last decade and offer significant therapeutic potential now and in future.

References

1. Tan DS (2005) Diversity-oriented synthesis: exploring the intersections between chemistry and biology. Nat Chem Biol 1: 74-84.
2. Webb TR (2005) Current directions in the evolution of compound libraries. Curr Opin Drug Discov Devel 8: 303-308.
3. Irwin JJ (2006) How good is your screening library? Curr Opin Chem Biol 10: 352-356.
4. Falciani C, Lozzi L, Pini A, Bracci L (2005) Bioactive peptides from libraries. Chem Biol 12: 417-426.
5. Ulrich H, Trujillo CA, Nery AA, Alves JM, Majumder P, et al. (2006) DNA and RNA aptamers: from tools for basic research towards therapeutic applications. Comb Chem High Throughput Screen 9: 619-632.
6. Uttamchandani M, Wang J, Yao SQ (2006) Protein and small molecule microarrays: powerful tools for high-throughput proteomics. Mol Biosyst 2: 58-68.
7. Goldberg AL, Akopian TN, Kisselev AF, Lee DH, Rohrwild M (1997) New insights into the mechanisms and importance of the proteasome in intracellular protein degradation. Biol Chem 378: 131-140.
8. Hilt W, Wolf DH (1996) Proteasomes: destruction as a programme. Trends Biochem Sci 21: 96-102.
9. Horswill AR, Savinov SN, Benkovic SJ (2004) A systematic method for identifying small-molecule modulators of protein-protein interactions. Proc Natl Acad Sci U S A 101: 15591-15596.
10. Matsumoto S, Akashi H, Taira K (2006) Screening and determination of gene function using randomized ribozyme and siRNA libraries. Handb Exp Pharmacol : 197-221.
11. Tavassoli A, Benkovic SJ (2007) Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. Nat Protoc 2: 1126-1133.
12. Danho W, Swistok J, Khan W, Xin-Jie Chu, Adrian Cheung, et al. (2009) Opportunities and challenges of developing peptide drugs in the pharmaceutical industry. Adv Exp Med Biol. 611:467-469
13. Vlieghe P, Lisowski V, Martinez J, Khrestchatisky M (2010) Synthetic therapeutic peptides: science and market. Drug Discov Today 15: 40-56.
14. Ayoub M, Scheidegger D (2006) Peptide drugs, overcoming the challenges, a growing business. Chem. Today 24: 46-48.
15. Edwards CM, Cohen MA, Bloom SR (1999) Peptides as drugs. QJM 92: 1-4.
16. Reubi JC (2003) Peptide receptors as molecular targets for cancer diagnosis and therapy. Endocr Rev 24: 389-427.
17. Bellmann-Sickert K, Beck-Sickinger AG (2010) Peptide drugs to target G protein-coupled receptors. Trends Pharmacol Sci 31: 434-441.
18. Lax R (2010) The Future of Peptide Development in the Pharmaceutical Industry. Phar Manufacturing: The International Peptide Review. 10-15.
19. Khafagy el-S, Morishita M (2012) Oral biodrug delivery using cell-penetrating peptide. Adv Drug Deliv Rev 64: 531-539.
20. Merrifield RB (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. J Am Chem Soc 85: 2149-2154.
21. Gutte B, Merrifield RB (1971) The synthesis of ribonuclease A. J Biol Chem 246: 1922-1941.
22. Barany G, Kempe M (1997) In: CW Czarnik, SH DeWitt (eds). A Practical Guide to Combinatorial Chemistry. Washington, DC: American Chemical Society 51-97.
23. Merrifield RB (1984) The role of the support in solid-phase peptide-synthesis. Br Polym J 16:173-178.
24. iely JS, Hayes TK, Griffith MC, Pei Y (1997) In: CW Czarnik, SH DeWitt,eds. A Practical Guide to Combinatorial Chemistry. Washington, DC: American Chemical Society 99-122.

25. Barany G, Albericio F, Sole NA, Griffin GW, Kates SA, Hudson D (1993) In: Ch Schneider, AN Eberle, (eds). Peptides 1992: Proceedings of the twenty-Second European Peptide Symposium. Leiden, The Netherlnds: ESCOM Science Publishers 267-268.

26. Bayer E, Rapp W (1992) In: JM Harris, (ed). Poly (Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications. New York: Plenum 325-345.

27. Barlos K, Gatos D, Kutsogianni S, Papaphotiou G, Poulos C, et al. (1991) Solid phase synthesis of partially protected and free peptides containing disulphide bonds by simultaneous cysteine oxidation-release from 2-chlorotrityl resin. Int J Pept Protein Res 38: 562-568.

28. Goodman M, Zapf C, Rew Y (2001) New reagents, reactions, and peptidomimetics for drug design. Biopolymers 60: 229-245.

29. Chan WC and White PD (2000) Fmoc solid phase peptide synthesis: a practical approach. Oxford, Oxford University Press 341 p. ISBN: 0-19-963724-5.

30. Tam JP, Merrifield RB (1987) Strong acid deprotection of synthetic peptides: mechanisms and methods. In: Udenfriend S and Meienbofer J, (eds). The Peptides. Analysis, Synthesis, Biology Vol. 9 Special Methods in Peptide Synthesis. New York: Academic Press, 185-248.

31. Yajima H, Fujii N, Funakoshi S, Watanabe T, Murayama E, et al. (1988) New strategy for the chemical synthesis of proteins. Tetrahedron 44: 805- 819.

32. Albericio F (2000) Orthogonal protecting groups for N(alpha)-amino and C-terminal carboxyl functions in solid-phase peptide synthesis. Biopolymers 55: 123-139.

33. Posnett DN, McGrath H, Tam JP (1988) A novel method for producing anti-peptide antibodies. Production of site-specific antibodies to the T cell antigen receptor beta-chain. J Biol Chem 263: 1719-1725.

34. Tam JP (1988) Synthetic peptide vaccine design: synthesis and properties of a high-density multiple antigenic peptide system. Proc Natl Acad Sci U S A 85: 5409-5413.

35. Tam JP, Zavala F (1989) Multiple antigen peptide. A novel approach to increase detection sensitivity of synthetic peptides in solid-phase immunoassays. J Immunol Methods 124: 53-61.

36. Tam JP, Clavijo P, Lu YA, Nussenzweig V, Nussenzweig R, et al. (1990) Incorporation of T and B epitopes of the circumsporozoite protein in a chemically defined synthetic vaccine against malaria. J Exp Med 171: 299-306.

37. Briand JP, Barin C, Van Regenmortel MH, Muller S (1992) Application and limitations of the multiple antigen peptide (MAP) system in the production and evaluation of anti-peptide and anti-protein antibodies. J Immunol Methods 156: 255-265.

38. Grillot D, Valmori D, Lambert PH, Corradin G, Del Giudice G (1993) Presentation of T-cell epitopes assembled as multiple-antigen peptides to murine and human T lymphocytes. Infect Immun 61: 3064-3067.

39. Christodoulides M, Heckels JE (1994) Immunization with a multiple antigen peptide containing defined B- and T-cell epitopes: production of bactericidal antibodies against group B Neisseria meningitidis. Microbiology 140 : 2951-2960.

40. Boykins RA, Joshi M, Syin C, Dhawan S, Nakhasi H (2000) Synthesis and construction of a novel multiple peptide conjugate system: strategy for a subunit vaccine design. Peptides 21: 9-17.

41. Zhao G, Lin Y, Du L, Guan J, Sun S, et al. (2010) An M2e-based multiple antigenic peptide vaccine protects mice from lethal challenge with divergent H5N1 influenza viruses. Virol J 7: 9.

42. Fujiki T, Tsuji A, Matsumoto SE, Yamashita M, Teruya K, et al. (2010) Generation of a human anti-tumor necrosis factor-Î± monoclonal antibody by *in vitro* immunization with a multiple antigen peptide. Biosci Biotechnol Biochem 74: 1836-1840.

43. Pau CP, Luo W, McDougal JS (2007) Chimeric multiple antigenic peptides for simultaneous detection of specific antibodies to HIV-1 groups M, N, O, and HIV-2. J Immunol Methods 318: 59-64.

44. Ali R, Kumar S, Naqvi RA, Sheikh IA, Rao DN (2013) Multiple antigen peptide consisting of B- and T-cell epitopes of F1 antigen of Y. pestis showed enhanced humoral and mucosal immune response in different strains of mice. Int Immunopharmacol 15: 97-105.

45. Hipkiss AR, Brownson C (2000) A possible new role for the anti-ageing peptide carnosine. Cell Mol Life Sci 57: 747-753.

46. Gill I, López-Fandiño R, Jorba X, Vulfson EN (1996) Biologically active peptides and enzymatic approaches to their production. Enzyme Microb Technol 18: 163-183.

47. Katzen F, Chang G, Kudlicki W (2005) The past, present and future of cell-free protein synthesis. Trends Biotechnol 23: 150-156.

48. Wright G, Carver A, Cottom D, Reeves D, Scott A, et al. (1991) High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. BioTechnology 9: 830 -834.

49. Cunningham C, Porter AJ (1997) Methods in Biotechnology, Vol. 3: Recombinant proteins from plants – production and isolation of clinically useful compounds. Totowa, N.J; Humana Press: 308 p. ISBN: 0-89-603390-2.

50. Vincent Du Vigneaud, Charlotte Ressler, College John M Swan, Carleton W Robets, Panayotis G Katsoyannis, et al. (1953) The synthesis of an octapeptide amide with the hormonal activity of oxytocin. Journal of the American Chemical Society 75: 4879-4880.

51. Feliú JA, De Mas C, Lópezsantín J (1995) Studies on papain action in the synthesis of Gly-Phe in two-liquid-phase media. Enzyme and Microbial Technology 17: 882-887.

52. Walsh G (2005) Therapeutic insulins and their large-scale manufacture. Appl Microbiol Biotechnol 67: 151-159.

53. Zhang L, Falla T, Wu M, Fidai S, Burian J, et al. (1998) Determinants of recombinant production of antimicrobial cationic peptides and creation of peptide variants in bacteria. Biochem Biophys Res Commun 247: 674-680.

54. Piers KL, Brown MH, Hancock RE (1993) Recombinant DNA procedures for producing small antimicrobial cationic peptides in bacteria. Gene 134: 7-13.

55. Hellers M, Gunne H, Steiner H (1991) Expression of post-translational processing of preprocecropin A using a baculovirus vector. Eur J Biochem 199: 435-439.

56. Sharma A, Khoury-Christianson AM, White SP, Dhanjal NK, Huang W, et al. (1994) High-efficiency synthesis of human alpha-endorphin and magainin in the erythrocytes of transgenic mice: a production system for therapeutic peptides. Proc Natl Acad Sci U S A 91: 9337-9341.

57. Osusky M, Zhou G, Osuska L, Hancock RE, Kay WW, et al. (2000) Transgenic plants expressing cationic peptide chimeras exhibit broad-spectrum resistance to phytopathogens. Nat Biotechnol 18: 1162-1166.

58. Lee JH, Kim JH, Hwang SW, Lee WJ, Yoon HK, et al. (2000) High-level expression of antimicrobial peptide mediated by a fusion partner reinforcing formation of inclusion bodies. Biochem Biophys Res Commun 277: 575-580.

59. Noritomi H, Suzuki K, Kikuta M, Kato S (2009) Catalytic activity of α-chymotrypsin in enzymatic peptide synthesis in ionic liquids. Biochem. Eng. J. 47: 27-30.

60. Kumar D, Bhalla TC (2005) Microbial proteases in peptide synthesis: approaches and applications. Appl Microbiol Biotechnol 68: 726-736.

61. Gupta MN, Roy I (2004) Enzymes in organic media. Forms, functions and applications. Eur J Biochem 271: 2575-2583.

62. Kamat S, Beckman EJ, Russell AJ (1992) Role of diffusion in nonaqueous enzymology. 1. Theory. Enzyme Microb Technol 14: 265-271.

63. Mesiano AJ, Beckman EJ, Russell AJ (1999) Supercritical Biocatalysis. Chem Rev 99: 623-634.

64. Gill I, Vulfson E (1994) Enzymic catalysis in heterogeneous eutectic mixtures of substrates. Trends Biotechnol 12: 118-122.

65. Halling PJ, Eichhorn U, Kuhl P, Jakubke HD (1995) Thermodynamics of solid-to-solid conversion and application to enzymic peptide synthesis. Enzyme and Microbial Technology 17: 601-606.

66. Erbeldinger M, Xiongwei NI, Halling, Peter J (1998) Enzymatic synthesis with mainly undissolved substrates at very high concentrations. Enzyme and Microbial Technology 23:141-148.

67. Park S, Kazlauskas RJ (2003) Biocatalysis in ionic liquids - advantages beyond green technology. Curr Opin Biotechnol 14: 432-437.

68. van Rantwijk F, Madeira Lau R, Sheldon RA (2003) Biocatalytic transformations in ionic liquids. Trends Biotechnol 21: 131-138.

69. Lou W, Zong M, Wu H (2004) Enhanced activity, enantioselectivity and stability of papain in asymmetric hydrolysis of D,L-p-hydroxyphenylglycine methyl ester with ionic liquid. Biocatalysis and Biotransformation 22: 171-176.

70. Machado MF, Saraiva JM (2005) Thermal stability and activity regain of horseradish peroxidase in aqueous mixtures of imidazolium-based ionic liquids. Biotechnol Lett 27: 1233-1239.

71. Illanes A, Barberis SE (1994) Catálisis enzimática en fase orgánica. In: Biotecnología de Enzimas. Monografía 35. Serie Biología. Secretaría General de la Organización de los Estados Americanos, Washington, DC, USA. Ediciones Universitarias de Valparaíso, Valparaíso. 225-254.

72. Quiroga E, Navigatore Fonzo L, Barberis S (2000) Catálisis enzimática en fase no acuosa. Parte I. Sistema de reacción. Anuario Latinoamericano de Educación Química ALDEQ 13: 20-24.

73. Meng LP, Joshi R, Eckstein H (2006) Application of enzymes for the synthesis of the cholecystokinin pentapepticle (CCK-5). A contribution to Green chemistry. Chim. Oggi. 24: 50-53.

74. Sakina K, Kawazura K, Morihara K, Yajima H (1988) Enzymatic synthesis of cholecystokinin-octapeptide. Chem Pharm Bull (Tokyo) 36: 3915-3919.

75. Bordusa F (2002) Proteases in organic synthesis. Chem Rev 102: 4817-4868.

76. Homandberg GA, Mattis JA, Laskowski M Jr (1978) Synthesis of peptide bonds by proteinases. Addition of organic cosolvents shifts peptide bond equilibria toward synthesis. Biochemistry 17: 5220-5227.

77. Jakubke HD, Kuhl P, Konnecke A (1985) Basic principles of protease-catalized peptide bond formation. Angewandte Chemie – International Edition in English 24: 85-93.

78. Chen YX, Zhang XZ, Zheng K, Chen SM, Wang QC, et al. (1998) Protease-catalyzed synthesis of precursor dipeptides of RGD with reverse micelles. Enzyme and Microbial Technology 23: 243-248.

79. Monter B, Herzog B, Stehle P, Fürst P (1991) Kinetically controlled synthesis of dipeptides using ficin as biocatalyst. Biotechnol Appl Biochem 14: 183-191.

80. Murakami Y, Hirata A (1997) New method for analysis of enzymatic reactions: thermolysincatalyzed synthesis of aspartame precursor. Journal of Fermentation and Bioengineering 83: 38-42.

81. Liu P, Tian G, Lee K, Wong M, Ye Y (2002) Full enzymatic synthesis of a precursor of bioactive pentapeptide OGP (10-14) in organic solvents. Tetrahedron Letters 43: 2423-2425.

82. Nakanishi K, Takeuchi A, Matsuno R (1990) Long-term continuous synthesis of aspartame precursor in a column reactor with an immobilized thermolysin. Appl Microbiol Biotechnol 32: 633-636.

83. Shaw C (1994) Peptide purification by reverse-phase HPLC. Methods Mol Biol 32: 275-287.

84. Chen H, Horváth C (1995) High-speed high-performance liquid chromatography of peptides and proteins. J Chromatogr A 705: 3-20.

85. Kiyama S, Fujii N, Yajima H, Moriga M, Takagi A (1984) Studies on peptides. CXIII. Synthesis of the heptacosapeptide amide corresponding to the entire amino acid sequence of chicken secretin. Int J Pept Protein Res 23: 174-186.

86. Filira F, Biondi L, Cavaggion F, Scolaro B, Rocchi R (1990) Synthesis of O-glycosylated tuftsins by utilizing threonine derivatives containing an unprotected monosaccharide moiety. Int J Pept Protein Res 36: 86-96.

87. Wolf FJ (1969) Separation Methods in Organic Chemistry and Biochemistry; Academic Press: New York, 1969.

88. Yamashiro D, Li CH (1978) Total Synthesis of ovine beta-lipotropin by the solid phase method. J.Am.Chem.Soc., 100: 5174-5179.

89. William S. Hancock (editor). (1995) New Methods in Peptide Mapping for the Characterization of Proteins. Boca Raton: CRC Press.

90. Swartz M, Murphy B (2004) Ultra performance liquid chromatography: Tomorrow's HPLC technology today. Lab Plus International 18: 6-9.

91. Winter D, Pipkorn R, Lehmann WD (2009) Separation of peptide isomers and conformers by ultra performance liquid chromatography. J Sep Sci 32: 1111-1119.

92. Jin Y, Song Y, Zhu X, Zhou D, Chen C, et al. (2012) Goblet cell-targeting nanoparticles for oral insulin delivery and the influence of mucus on insulin transport. Biomaterials 33: 1573-1582.

93. Kurzrock R, Gabrail N, Chandhasin C, Moulder S, Smith C, et al. (2012) Safety, pharmacokinetics, and activity of GRN1005, a novel conjugate of angiopep-2, a peptide facilitating brain penetration, and paclitaxel, in patients with advanced solid tumors. Mol Cancer Ther 11: 308-316.

94. Kapoor DN, Katare OP, Dhawan S (2012) In situ forming implant for controlled delivery of an anti-HIV fusion inhibitor. Int J Pharm 426: 132-143.

95. Dai W, Yang T, Wang Y, Wang X, Wang J, et al. (2012) Peptide PHSCNK as an integrin Î±5Î²1 antagonist targets stealth liposomes to integrin-overexpressing melanoma. Nanomedicine 8: 1152-1161.

96. Däpp S, García Garayoa E, Maes V, Brans L, Tourwé DA, et al. (2011) PEGylation of (99m)Tc-labeled bombesin analogues improves their pharmacokinetic properties. Nucl Med Biol 38: 997-1009.

97. Zhao Y, Li J, Yu H, Wang G, Liu W (2012) Synthesis and characterization of a novel polydepsipeptide contained tri-block copolymer (mPEG-PLLA-PMMD) as self-assembly micelle delivery system for paclitaxel. Int J Pharm 430: 282-291.

98. Cai L, Wang X, Wang W, Qiu N, Wen J, et al. (2012) Peptide ligand and PEG-mediated long-circulating liposome targeted to FGFR overexpressing tumor in vivo. Int J Nanomedicine 7: 4499-4510.

99. Mitra R, Ganesh KN (2012) Aminomethylene peptide nucleic acid (am-PNA): synthesis, regio-/stereospecific DNA binding, and differential cell uptake of (Î±/Î³,R/S)am-PNA analogues. J Org Chem 77: 5696-5704.

100. Jie LY, Cai LL, Wang LJ, Ying XY, Yu RS, et al. (2012) Actively-targeted LTVSPWY peptide-modified magnetic nanoparticles for tumor imaging. Int J Nanomedicine 7: 3981-3989.

101. Alam S, Panda JJ, Chauhan VS (2012) Novel dipeptide nanoparticles for effective curcumin delivery. Int J Nanomedicine 7: 4207-4222.

102. Jabbari E, Yang X, Moeinzadeh S, He X (2013) Drug release kinetics, cell uptake, and tumor toxicity of hybrid VVVVVVKK peptide-assembled

polylactide nanoparticles. Eur J Pharm Biopharm 84: 49-62.

103. SlaninovÃ¡ J, MlsovÃ¡ V, KroupovÃ¡ H, AlÃ¡n L, TÅ¯movÃ¡ T, et al. (2012) Toxicity study of antimicrobial peptides from wild bee venom and their analogs toward mammalian normal and cancer cells. Peptides 33: 18-26.

104. Rossi DC, Muñoz JE, Carvalho DD, Belmonte R, Faintuch B, et al. (2012) Therapeutic use of a cationic antimicrobial peptide from the spider *Acanthoscurria gomesiana* in the control of experimental candidiasis. BMC Microbiol 12: 28.

105. Drin G, Cottin S, Blanc E, Rees AR, Temsamani J (2003) Studies on the internalization mechanism of cationic cell-penetrating peptides. J Biol Chem 278: 31192-31201.

106. Rousselle C, Clair P, Smirnova M, Kolesnikov Y, Pasternak GW, et al. (2003) Improved brain uptake and pharmacological activity of dalargin using a peptide-vector-mediated strategy. J Pharmacol Exp Ther 306: 371-376.

107. Adenot M, Merida P, Lahana R (2007) Applications of a blood-brain barrier technology platform to predict CNS penetration of various chemotherapeutic agents. 2. Cationic peptide vectors for brain delivery. Chemotherapy 53: 73-76.

108. Tian XH, Wei F, Wang TX, Wang P, Lin XN, et al. (2012) *In vitro* and *in vivo* studies on gelatin-siloxane nanoparticles conjugated with SynB peptide to increase drug delivery to the brain. Int J Nanomedicine 7: 1031-1041.

109. Li Y, Zheng X, Cao Z, Xu W, Zhang J, et al. (2012) Self-assembled peptide (CADY-1) improved the clinical application of doxorubicin. Int J Pharm 434: 209-214.

110. Du YZ, Cai LL, Liu P, You J, Yuan H, et al. (2012) Tumor cells-specific targeting delivery achieved by A54 peptide functionalized polymeric micelles. Biomaterials 33: 8858-8867.

111. Wu Y, Sadatmousavi P, Wang R, Lu S, Yuan YF, et al. (2012) Self-assembling peptide-based nanoparticles enhance anticancer effect of ellipticine *in vitro* and in vivo. Int J Nanomedicine 7: 3221-3233.

112. Khaksa G, D'Souza R, Lewis S, Udupa N (2000) Pharmacokinetic study of niosome encapsulated insulin. Indian J Exp Biol 38: 901-905.

113. Reubi JC (2003) Peptide receptors as molecular targets for cancer diagnosis and therapy. Endocr Rev 24: 389-427.

114. SimÃµes S, Moreira JN, Fonseca C, DÃ¼zgÃ¼neÅŸ N, de Lima MC (2004) On the formulation of pH-sensitive liposomes with long circulation times. Adv Drug Deliv Rev 56: 947-965.

115. Samad A, Sultana Y, Aqil M (2007) Liposomal drug delivery systems: an update review. Curr Drug Deliv 4: 297-305.

116. Irache JM, Salman HH, Gamazo C, Espuelas S (2008) Mannose-targeted systems for the delivery of therapeutics. Expert Opin Drug Deliv 5: 703-724.

117. Grant M, Leone-Bay A (2012) Peptide therapeutics: it's all in the delivery. Ther Deliv 3: 981-996.

118. Yamada A, Sasada T, Noguchi M, Itoh K (2013) Next-generation peptide vaccines for advanced cancer. Cancer Sci 104: 15-21.

119. Mahato RI, Narang AS, Thoma L, Miller DD (2003) Emerging trends in oral delivery of peptide and protein drugs. Crit Rev Ther Drug Carrier Syst 20: 153-214.

120. Khafagy el-S, Morishita M (2012) Oral biodrug delivery using cell-penetrating peptide. Adv Drug Deliv Rev 64: 531-539.

121. He XH, Shaw PC, Tam SC (1999) Reducing the immunogenicity and improving the *in vivo* activity of trichosanthin by site-directed pegylation. Life Sci 65: 355-368.

122. Grace MJ, Lee S, Bradshaw S, Chapman J, Spond J, et al. (2005) Site of pegylation and polyethylene glycol molecule size attenuate interferon-alpha antiviral and antiproliferative activities through the JAK/STAT signaling pathway. J Biol Chem 280: 6327-6336.

123. Sun LC, Coy DH (2011) Somatostatin receptor-targeted anti-cancer therapy. Curr Drug Deliv 8: 2-10.

124. Ahn JM, Boyle NA, MacDonald MT, Janda KD (2002) Peptidomimetics and peptide backbone modifications. Mini Rev Med Chem 2: 463-473.

125. Lozano JM, Espejo F, Ocampo M, Salazar LM, Tovar D, et al. (2004) Mapping the anatomy of a Plasmodium falciparum MSP-1 epitope using pseudopeptide-induced mono- and polyclonal antibodies and CD and NMR conformation analysis. J Struct Biol 148: 110-122.

126. Våbenø J, Lejon T, Nielsen CU, Steffansen B, Chen W, et al. (2004) Phe-Gly dipeptidomimetics designed for the di-/tripeptide transporters PEPT1 and PEPT2: synthesis and biological investigations. J Med Chem 47: 1060-1069.

127. Albericio F, Burger K, Ruíz-Rodríguez J, Spengler J (2005) A new strategy for solid-phase depsipeptide synthesis using recoverable building blocks. Org Lett 7: 597-600.

128. Cruz LJ, Insua MM, Baz JP, Trujillo M, Rodriguez-Mias RA, et al. (2006) IB-01212, a new cytotoxic cyclodepsipeptide isolated from the marine fungus Clonostachys sp. ESNA-A009. J Org Chem 71: 3335-3338.

129. Strömstedt AA, Pasupuleti M, Schmidtchen A, Malmsten M (2009) Evaluation of strategies for improving proteolytic resistance of antimicrobial peptides by using variants of EFK17, an internal segment of LL-37. Antimicrob Agents Chemother 53: 593-602.

130. Wrighton NC, Farrell FX, Chang R, Kashyap AK, Barbone FP, et al. (1996) Small peptides as potent mimetics of the protein hormone erythropoietin. Science 273: 458-464.

131. Wirsching F, Keller M, Hildmann C, Riester D, Schwienhorst A (2003) Directed evolution towards protease-resistant hirudin variants. Mol Genet Metab 80: 451-462.

132. Odegrip R, Coomber D, Eldridge B, Hederer R, Kuhlman PA, et al. (2004) CIS display: *In vitro* selection of peptides from libraries of protein-DNA complexes. Proc Natl Acad Sci U S A 101: 2806-2810.

133. Eldridge W, Fitzgerald K, Cooley N, McGregor D (2006) Peptide stabilizer compounds and screening methods. Patent Application WO2006097748.

134. Woodburn KW, Wilson SD, Fong KL, Schatz PJ, Spainhour CB, et al. (2009) Chronic pharmacological and safety evaluation of Hematide, a PEGylated peptidic erythropoiesis-stimulating agent, in rodents. Basic Clin Pharmacol Toxicol 104: 155-163.

135. Ehrlich A, Heyne HU, Winter R, Beyermann M, Haber H, et al. (1996) Cyclization of all-L-Pentapeptides by Means of 1-Hydroxy-7-azabenzotriazole-Derived Uronium and Phosphonium Reagents. J Org Chem 61: 8831-8838.

136. Lioy E, Suarez J, Guzman F, Siegrist S, Pluschke G, et al. (2001) Manuel E. Synthesis, biological, and immunological properties of cyclic peptides from Plasmodium falciparum merozoite surface protein-1. Angewandte Chemie - International Edition in English 40: 2631-2635.

137. Gururaja TL, Narasimhamurthy S, Payan DG, Anderson DC (2000) A novel artificial loop scaffold for the noncovalent constraint of peptides. Chem Biol 7: 515-527.

138. Pakkala M, Hekim C, Soininen P, Leinonen J, Koistinen H, et al. (2007) Activity and stability of human kallikrein-2-specific linear and cyclic peptide inhibitors. J Pept Sci 13: 348-353.

139. Fernandez-Lopez S, Kim HS, Choi EC, Delgado M, Granja JR, et al. (2001) Antibacterial agents based on the cyclic D,L-alpha-peptide architecture. Nature 412: 452-455.

140. Dartois V, Sanchez-Quesada J, Cabezas E, Chi E, Dubbelde C, et al. (2005) Systemic antibacterial activity of novel synthetic cyclic peptides.

Antimicrob Agents Chemother 49: 3302-3310.

141. Colas P, Cohen B, Jessen T, Grishina I, McCoy J, et al. (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. Nature 380: 548-550.

142. von Moos E, Ben RN (2005) Recent advances in the synthesis of C-linked glycoconjugates. Curr Top Med Chem 5: 1351-1361.

143. Engel JB, Schally AV, Halmos G, Baker B, Nagy A, et al. (2005) Targeted cytotoxic bombesin analog AN-215 effectively inhibits experimental human breast cancers with a low induction of multi-drug resistance proteins. Endocr Relat Cancer 12: 999-1009.

144. Keller G, Schally AV, Nagy A, Baker B, Halmos G, et al. (2006) Effective therapy of experimental human malignant melanomas with a targeted cytotoxic somatostatin analogue without induction of multi-drug resistance proteins. Int J Oncol 28: 1507-1513.

145. Sun LC, Coy DH (2011) Somatostatin receptor-targeted anti-cancer therapy. Curr Drug Deliv 8: 2-10.

146. Derossi D, Joliot AH, Chassaing G, Prochiantz A (1994) The third helix of the Antennapedia homeodomain translocates through biological membranes. J Biol Chem 269: 10444-10450.

147. Vivès E, Brodin P, Lebleu B (1997) A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. J Biol Chem 272: 16010-16017.

148. Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, et al. (2000) The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. Proc Natl Acad Sci U S A 97: 13003-13008.

149. Madani F, Lindberg S, Langel U, Futaki S, Gräslund A (2011) Mechanisms of cellular uptake of cell-penetrating peptides. J Biophys 2011: 414729.

150. Engel JB, Schally AV, Halmos G, Baker B, Nagy A, et al. (2005) Targeted cytotoxic bombesin analog AN-215 effectively inhibits experimental human breast cancers with a low induction of multi-drug resistance proteins. Endocr Relat Cancer 12: 999-1009.

151. Keller G, Schally AV, Nagy A, Baker B, Halmos G, et al. (2006) Effective therapy of experimental human malignant melanomas with a targeted cytotoxic somatostatin analogue without induction of multi-drug resistance proteins. Int J Oncol 28: 1507-1513.

152. Silva AL, Rosalia RA, Sazak A, Carstens MG, Ossendorp F, et al. (2013) Optimization of encapsulation of a synthetic long peptide in PLGA nanoparticles: low-burst release is crucial for efficient CD8(+) T cell activation. Eur J Pharm Biopharm 83: 338-345.

153. Dombu CY, Betbeder D (2013) Airway delivery of peptides and proteins using nanoparticles. Biomaterials 34: 516-525.

154. Marx V (2005) Watching Peptide Drugs Grow up. C&EN 83: 17-24.

Advances in Protein Chemistry

Chapter: Protein Detection Techniques

Edited by: Ghulam Md Ashraf and Ishfaq Ahmed Sheikh

Published by **OMICS Group eBooks**

731 Gull Ave, Foster City. CA 94404, USA

Cover OMICS Group Design team

First published November, 2013

A free online edition of this book is available at www.esciencecentral.org/ebooks

Additional hard copies can be obtained from orders @ www.esciencecentral.org/ebooks

Protein Detection Techniques

Dennis F Thekkudan*

Department of Chemistry, Virginia Commonwealth University, Richmond VA 23284-2006, USA

***Corresponding author:** Dennis F Thekkudan, Department of Chemistry, Virginia Commonwealth University, Richmond VA 23284-2006, USA, Tel: (757) 439-3752; E-mail: DennisT44@gmail.com

Abstract

This review provides a synopsis of some of the more contemporary techniques of studying proteins and peptides. Bottom-up proteomics, as opposed to top-down proteomics, is becoming more common place in protein sequencing. Liquid chromatography-mass spectrometry (LC-MS) is being used to resolve compounds of interest chromatographically before investigating them with the mass spectrometry (MS) detector to take full advantage of its sensitivity. Ionization procedures are also utilized to clean up the sample matrix and labeling methodologies are used to help in quantifying these compounds. Matrix adsorption laser desorption-ionization/imaging mass spectrometry (MALDI-IMS) is a novel technique to assist in envisioning biological changes within a given tissue. Finally, atomic force microscopy (AFM) and single-molecule force spectroscopy (SMFS) were studied to offer understanding into the changes in secondary structure of proteins. It is obvious that the techniques used to study proteins and peptides are diverse; nevertheless, they are now becoming more conventional in modern analytical investigations. Knowledge of how these systems work will allow us to interpret the sequence of peptides and understand the function of proteins in biological systems.

Introduction

The broad analysis of proteins using a mass spectrometry (MS) as a detector along with a separation technique can help us gain more information about the structure and function of proteins. The quick and efficient detection of proteins via MS has been advanced by the creation of soft-ionization methods like matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), especially when coupled to liquid chromatography (LC). A thorough analysis of a protein involves examining how a peptide fragments in an MS detector. LC in conjunction with MS provides an abundance of data, while LC with tandem MS (MS/MS) provides even more complexity. Programs and tools to sift through the mass amounts of data are needed to interpret this data. Using chromatography to separate different classes of proteins from each other is one example of the top-down approach. However, this approach is now falling out of favor [1].

Top-down proteomics involves performing a chromatographic separation of complete proteins and analyzing them by using high-resolution MS or tandem MS. High-resolution enables the analyst to examine masses typically down to the thousandth of a Dalton. As a result, the analyst can single out a peptide based on its exact mass. An example of a high-resolution mass spectrometer is an ion cyclotron resonance mass spectrometer [2]. This MS has mass resolution on the order of 100,000, which is the greatest among MS detectors. MS/MS is used to detect molecular ions and then fragment them down to smaller ions. Based on the fragmentation patterns of the initial molecular ions, an analyst can deduce the structure of a peptide. The primary advantage to top-down proteomics is that it enables the detection of post-translational modifications. However, the downside of the top-down approach is its lack of sensitivity [3].

The technique typically performed now is bottom-up proteomics where the protein is enzymatically cleaved prior to analysis because the MS instrument is capable of characterizing multiple protein fragments. Trypsin is typically used to cleave the protein prior to analysis. This can provide anywhere from 10 to 100 fragments prior to any fragmentation provided by the tandem MS. Typically, the MS or tandem MS data are interpreted by searching sophisticated databases of common peptide and protein fragments. Most software packages that come with an MS allow you to reshape the data into either a tab delimited values or into a comma separated value file. The files can be imported into a peptide searching database such as OMSSA (http://pubchem.ncbi.nlm.nih.gov//omssa) from NIST, Bethesda, MD, USA or Mascot (http://matrixscience.com) from Matrix Science, London, UK, which accepts most raw data files from the MS software itself. It should be noted that different search algorithms provide distinct methods for ranking potential matches and can provide different interpretations of the data [1].

A new approach within the scope of the bottom-up approach is referred to as multidimensional protein identification technology or MudPIT. It is essentially a two-dimensional LC method in which solutions made out of digested peptides are separated by using a cation exchange column as the first column and a reverse-phase LC column as the second column. According to Yateset al., this separation process is performed on-line and then sprayed into the tandem MS detector [4]. The spectra are subsequently searched using an algorithm. The performance of bottom-up proteomics via LC-MS/MS is much more efficient and less labor intensive than the analysis of scraped spots from two-dimensional gel electrophoresis. The primary downside of the bottom-up approach is that it does not allow for the analysis of post-translational modifications [3].

Chromatographic Techniques

Size exclusion chromatography or SEC is a mode of separation where the largest and heaviest compounds elute first while the smaller and lighter compounds elute last because the smaller compounds have to traverse through a maze of pores within the chromatographic column [5]. One disadvantage of SEC is that it provides low peak capacity, i.e., typically on the order of 10. Peak capacity refers to the amount of resolved peaks over a specified retention time period within a chromatogram [6]. Sieving gels, which also separates based on molecular weight, provide a peak capacity of approximately 50. The peaks obtained from SEC are wider than the windows into which these peaks must fit. It is for the above reasons that SEC is not the ideal chromatographic method for proteomics [5].

On the other hand, high-performance liquid chromatographic (HPLC) separations, specifically reverse-phase liquid chromatography (RPLC), provide an opportunity to isolate proteins based on their hydrophobicity. Reverse phase separations typically involve using a column with a stationary phase that is hydrophobic. This stationary phase is generated by binding a long chain hydrocarbon with a surface made of silica via a siloxane bond. The hydrocarbon layer is approximately 10 Å thick, which enables efficient mass transport to the mobile phase. Gradient mobile phases are typically employed where the initial composition is highly aqueous and the ending composition is highly organic using methanol, acetonitrile or tetrahydrofuran. In the context of peptides and proteins, the most hydrophilic compounds elute first and the most hydrophobic compounds elute last. As a result, RPLC provides an effective complement to SEC, which separates peptides and proteins on the basis of molecular weight [5].

Chromatographic columns, however, are not an ideal facilitator for two-dimensional separations. Two dimensional separations are achieved by placing the two columns in sequence with the first separation occurring on the order of minutes and the second separation taking place on the order of seconds [7]. Two-dimensional liquid chromatography is more complex than preparing a two-dimensional thin-layer chromatography plate. A three-dimensional peptide separation was achieved by Moore and Jorgenson where ovalbumin that has undergone a tryptic digest procedure was separated by placing an SEC column, an RPLC column and a capillary zone electrophoresis (CZE) setup in order [8]. The peptides were satisfactorily resolved from each other. The SEC separation, in this case, lasted for 80 minutes, the RPLC separation was 90 seconds, and CZE separation was 27 seconds. The process of sampling and timing is not of concern in two-dimensional gel electrophoresis as it is in multidimensional chromatography. In addition, care must be taken when selecting mobile phases that are compatible with the neighboring separation mode. Hu et al. attained fast two-dimensional separations when performing proteomics on individual cells by using two capillary columns; specifically, micellar electrokinetic chromatography and CZE using separation modes of hydrophobicity and charge, respectively [9]. High sensitivity was attained using this technique, which may be a precursor for studying biomarkers and disease mechanisms [5].

Ionization Considerations

One constraint when performing LC-MS is the flow rate provided by the LC. The design of the MS detector and the efficiency of the vacuum pump are the primary reasons why we have to control the flow rate of the LC. For the ESI, the flow rate should typically be set from 20 to 1,000 µL per minute. Exceptions include high aqueous mobile phases in which the flow rate can be set to 200 to 500 µL per minute. If it is mandatory that your mobile phase must pump at a higher flow rate, splitting the flow to the MS detector is needed; otherwise, the ESI is overwhelmed and it cannot effectively produce gas-phase ions. For APCI, flow rates higher than 1,000 µL per minute are mandatory for sample ionization. If it required that a lower flow rate is needed to gain separation, a makeup flow, which is an additional flow of solvent or mobile phase, is pumped in after the LC column to assist in the ionization process. Finally, the capillary tubing from the end of the LC column to the MS detector should be as short and as narrow as possible to improve chromatographic peak resolution and to prevent band broadening [10].

LC-MS initially used one of two ionization methods: atmospheric pressure chemical ionization (APCI) and ESI [3]. These atmospheric pressure ionization (API) methods requires that the mobile phase flows through a narrow capillary to be subsequently ionized into a spray of gas-phase molecules due to the evaporation of solvent under atmospheric pressure. These gas-phase ions then enter an orifice or a pinhole into the MS detector, which is under high-vacuum conditions. The MS then detects these gas phase ions. There are quite a few constraints that need to be known when operating an LC-MS. The primary constraint is that non-volatile buffers can quench the MS signal. These buffers include salts such as sodium chloride or phosphate ions that can impede with the capillary and physically block the opening into the detector. Protonation of chemical compounds is a technique used to enhance the MS signal. This practice, however, is only useful when cations are absent from the sample. When salts or phosphate ions are present in a chemical system, the mass spectrum detects chemical complexes with multiple cations, i.e., [M + Na]+, [M+2Na]2+, etc. As a result, the sensitivity of the compound you wish to detect is quenched. Typically, either ammonium formate or ammonium acetate as buffer salts is used in low concentrations to enhance a MS signal. These two buffer salts act to create weak cation compounds that maximize the sensitivity of the MS detector [10].

ESI involves a solution moving through an electrostatically charged needle to create droplets with either a positive or negative charge. As a result, the droplets have a charge (positive or negative) that is the same of that of the needle (positive or negative, respectively). Rayleigh dissociation is described by the process of a droplet shrinking and bursting into smaller droplets due to the excess charge generated. The ion evaporation model best describes small molecules including peptides where the strong electric fields force the ions to be released from the droplet thereby producing gas-phase ions. The charge residue model is an alternate mechanism that best describes large molecules including proteins where solvent making up the charged droplet evaporates to produce the gas-phase ion. No matter what the mechanism is, the result is gas-phase ions. ESI is ultimately advantageous because it is easy to interface to LC, it enables the ionization of a large variety of molecules and it facilitates the sensitivity of detection [3].

APCI produces gas-phase ions by using a heated nebulizer for the mobile phase. Specifically, the ionization takes place in plasma created by a corona discharge where there is a transfer of a proton from the eluent to the analyte. APCI is limited to smaller molecules because it requires the vaporization of the analyte before ionization. However, the advantage of APCI over ESI is the fact that the sample matrix does not have an effect on the ability of APCI to produce gas-phase ions. There is also a process called atmospheric pressure photoionization in which the corona discharge is substituted by a xenon arc lamp [3].

Stable Isotope Labeling

There are three approaches to performing quantification using stable isotope labeling: isobaric tagging for relative and absolute quantitation or iTRAQ, isotope-coded affinity tags or ICAT, and stable isotope labeling by amino acids in cell culture or SILAC [3]. iTRAQ enables analysis of eight samples at the same time and is a newer labeling methodology. It involves a compound that inserts a shared change in mass to every free amino group. Specifically, the lysine residues and N-terminus are derivatized with this compound. As a result, fragment ions emerge ranging from 114 m/z to 121 m/z depending on the compound used for derivatization [11]. Different iTRAQ compounds act to derivatize samples undergoing comparison. This derivatization helps to quantify changes within the sample [3]. The first use of labeling for quantitative proteomics was ICAT, which was presented by Gygi and coworkers in 1999 [12]. The compounds used for ICAT introduce stable isotopes into proteins by derivatizing certain amino acids such as cysteine within mixtures of proteins with a compound containing biotin. Alternate types of the ICAT compound can be used to derivatize samples. The samples treated with ICAT compounds are combined together with those samples that are untreated before they are assayed. Exact quantification of proteins within complex mixtures and identification of a protein sequence are the results of combining tandem MS with ICAT. This

technique has been applied to the investigation of protein secretion within bodily fluids, constituents of protein complexes, changes in proteins within subcellular fractions and overall changes in the expression of proteins [3].

Inserting isotopes within proteins by developing mammalian cells in media using a stable isotope of an amino acid describes SILAC [13]. Quantification of protein expression is determined by comparing cells produced using SILAC against cells untreated by SILAC. Both of these samples can be analyzed within the same assay because of the m/z change triggered by the stable isotope label. SILAC is, however, constrained to applications of tissue culture [3]. A specific example of SILAC used for protein quantification is metabolic labeling using ^{15}N *in vivo*. Amino acids labeled with ^{15}N isotopes are used in the synthesis of a protein. Both the ^{14}N and the ^{15}N protein isotopes are mixed before sample preparation to act as an internal standard for quantification especially when using LC-MS. In LC-MS, an internal standard that is separated from the compound of interest can ionize at different rates using the soft ionization techniques of ESI or MALDI, especially if using a gradient elution method. Using an isotopically labeled compound enables both the compound of interest and the internal standard to be measured at the same ionization conditions [3].

A primary issue when performing proteomics is quantification. There are two approaches for quantification in proteomics: label-free methods and label methods. Label methods involve using stable isotope labels to act as an internal standard without quenching the signal of the analyte. As a result, the reproducibility of the data is enhanced. In contrast, there are some label methods that are cheaper and easier to use along with the fact that they provide acceptable reproducibility. By using a linear ion trap MS with a label-free methodology, Higgs et al. was able to achieve less than 10 % relative standard deviation (RSD) for their bottom-up proteomic application [14]. Differential MS is another label-free methodology used by Wiener and coworkers in which ions that demonstrate differences in intensity that are statistically significant are additionally probed by tandem MS for sequencing [15]. Since the results that show a statistical difference are targeted for identification, this methodology enables efficient calculations [3].

Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry

The primary advantage of MS over other analytical detectors is its specificity to molecules. The ionization techniques discussed earlier described preparing a liquid sample of MS analysis. Utilization of matrix-assisted laser desorption/ionization or MALDI enables the scientist to study intact tissue and the distribution of the proteins within this tissue. This ionization technique needs to be matched with imaging mass spectrometry or IMS to analyze tissue specimens. The direct analysis of tissue specimens entails minimal effort and time in sample preparation and enables direct comparison between different biological species or samples [16].

With a lateral image resolution 300 to 500 Å, IMS can detect a variety of molecules such as proteins, fatty acids, and pharmaceutics from sections of frozen fresh tissue. The analyst first cuts sections of the tissue down to 100 to 150 Å thick and then mounts them on target plates where they can equilibrate to room temperature. After these steps, the analyst applies an energy-absorbing compound or a matrix to the tissue section. A UV laser then ablates an area of the tissue that is approximately 500 Å wide. This process results in molecular ions that are detected via the MS detector. Each area that is ablated in the section corresponds to a single MS. Extraction of the relative abundances of an ion for a single m/z ratio can then provide a contour plot with the ion intensities projecting outward and the location of the section given on the x and y axes. It is critical to deposit the energy-absorbing compound in a homogeneous manner to obtain a high-resolution image. Either using a robot to spray the matrix with a continuous coating or automatically printing groupings of droplets are techniques to accomplish this task. MALDI MS then probes each pixel coordinate or micro spot. Provided that signals are above the limit of quantification, a contour plot can be obtained at any m/z ratio. Ultimately, MALDI is operational for identification and quantification of both infected and healthy biological tissue. In addition, it can provide insight into describing changes in a biological system. Applications of this technique include reaction of a tissue to administration of a drug, reconstructions of a biological organ, and three-dimensional images of the brain and its proteins [16].

The primary advantage of IMS is that it can automatically image dozens upon dozens of compounds in tissue sections with targeting compounds, antibodies or even prior knowledge of the tissue being analyzed. IMS allows for imaging of proteolytic catalysis or post-translational modifications unlike most bottom-up proteomic assays. While not effective for proteomic analysis, secondary ionization mass spectrometry or SIMS enables high-resolution imaging, on the order of 500 to 1000 Å, for molecules less than 1000 m/z down to elemental species [16].

Imaging technology

There are two general methodologies to analyzing biological tissue via IMS: imaging and profiling. Probing distinct regions (200 to 1000 µm in diameter) of tissue slices and performing computational analysis of the resulting mass spectra (typically, 5 to 20 in quantity) is a general characterization of the profiling methodology. This method compares two different tissue specimens or two significant regions, i.e., diseased vs. healthy, within a tissue. As a result, it is not necessary to have a high degree of spatial resolution [16].

Imaging involves analyzing the entirety of the tissue section by systematically sampling spots on the tissue. These spots, or raster, are a constant distance away from each other both vertically and horizontally. This distance defines the image resolution. The ion abundance, perhaps of a certain m/z, is plotted against the xy coordinates. This plot can then be used to assess the differences of protein localization among samples [16].

To reduce degradation of proteins via proteolysis and to preserve morphology, the tissue sections must be submerged in liquid N2 immediately. Typically, the tissue is sliced into 100 Å to 120 Å sections within a cryostat and subsequently mounted on a sample plate that is constructed of stainless steel or coated with gold to conduct electricity. Ethyl alcohol is used to wash the tissue to help mount it and to rinse away excess salts and lipids thereby preventing ion suppression. Another option is to conduct protocols that stain tissue with compounds that specifically facilitate IMS detection [16].

The use of a matrix that absorbs energy is necessary for MALDI IMS. Typically, it is an organic molecule of small size that crystallizes simultaneously with your compound of interest on the surface of the tissue. It absorbs the energy from a laser enabling your compound of interest to desorb and then ionize. Typical matrices used for MALDI are α-cyano-4-hydroxy-cinnamic acid (HCCA), 2, 5-dihydroxybenzoic acid (DHB), and 3, 5-dimethoxy-4-hydroxy-cinnamic acid (SA). Typical solvents used for analysis include 50:50 (v/v) water/ethyl alcohols or 50:50 (v/v) water/acetonitrile. HCCA and DHB are typically used for peptides and smaller compounds, while SA is typically used for protein molecules [16].

Deposition of the matrix should be homogeneous to facilitate high-resolution imaging and to preclude substantial lateral migration

of the compounds of interest. High-resolution imaging is typically accomplished by using a homogenous spray coating or a spotted array. The best spatial resolution is achieved by a spray coating that is homogeneous while improved spectral quality and precision is achieved by using an array that is densely spotted. On the other hand, a heterogeneous coating provides locations on the specimen that are either dilute or concentrated for ablation. Crystal formation, as a result, is random and images are both highly pixilated and poor in quality. Robotic spotting instruments are commercially accessible and use a variety of techniques such as capillary deposition, ink-jet printer, piezoelectric and acoustic methods. Robotic spray-coating instruments are also available that use either a thermally assisted spray or a mist-nebulizer [16].

The best type of MS detector for proteomic analysis of biological tissue is the time-of-flight or TOF MS. The TOF detector complements the pulsed laser used by MALDI. In addition, it is easy to maintain, has a simple design, is capable of detecting multiple m/z ratios, is efficient in ion-transmission, and has a theoretically unlimited dynamic range [16]. Ions generated during the MALDI process are accelerated through the field-free drift tube of the TOF detector. The ions collect at the multichannel plate detector where the small m/z ions arrive first and the large m/z ions arrive last. If properly calibrated, then the arrival time is converted to m/z ratio via

$$\frac{m}{z} = \frac{2Vt^2}{L^2}$$

Where V is the voltage potential applied to accelerate ions, t is the time the ion arrives at the detector, and L is the length of the field-free drift tube [2].

MALDI/IMS quantification

To quantify proteins or peptides using MALDI IMS, the precision among pixels should be good. In other words, two neighboring pixels with the same protein concentrations should provide the same mass spectra. These spectra typically do not differ from each other by more than 15%. Some of the items to consider if there is a significant difference include extraction efficiency of the MALDI, ionization efficiency of a particular compound, ion-suppression effects and the effect of post-acquisition processing. In particular, sample preparation and matrix application are the two techniques that need to be mastered to obtain good precision among pixels. A constant laser power and voltage should also be maintained within the assay to achieve maximum precision. To prevent variation among operators, robotics is typically employed. It is important to note that even if these parameters are kept under control and high reproducibility is achieved, estimation of relative concentrations of two different proteins is still difficult to obtain by comparing peak areas or peak heights [16].

MS reproducibility should be accurately estimated when attempting to correlate MS signal intensities to values with biological meaning. Statistical analysis and data preprocessing are tools a scientist can use to probe the biological meaning from these data. Data preprocessing steps include background removal algorithms, signal intensity normalization and peak alignment [17]. The purpose of the normalization of signal intensity to the total ion current is to minimize variations in the sample preparations. As a result of these data preprocessing steps, a new ion image can be generated that better represents the biological system [16].

After the data preprocessing steps are completed, an average mass spectrum is either taken of each target area within a specimen or of each tissue specimen itself. Statistical analyses are then performed on each of these spectra. Principal component analysis (PCA) is a methodology used to rank the variance in the system to determine the number of significant components that are in the system [17]. Scores can be assigned to components based on how it changes its signal intensity relative to the standard deviation of multiple experiments. This methodology is called significance analysis of microarrays or SAM. The changes within multiple experiments help to estimate the fraction of components that are seen by chance (potential false positives) when scores are greater than a predetermined threshold. SAM specifically exposes components that demonstrate a significant change between two groups of tissues. A method that blindly groups samples based on their profiles is hierarchical clustering analysis or HCA. HCA specifically computes the dissimilarity between particular experiments [16]. Based on the dissimilarity values, a dendrogram is created, which clusters or groups similar components and depicts these similar groups in a graph that looks similar to a tournament bracket [17].

Surface Characterization

A specific instrument used for the characterization of proteins is the atomic force microscope (AFM). The AFM is made out of a cantilever with a tip where van der Waals forces are measured between the tip and the sample surface. These van der Waals forces cause the tip to be redirected, and this movement is optically detected by the instrument. Constant force applied to tip enables a strict up-and-down movement of the tip providing data on the topography of the sample surface. The flexibility of the AFM lies in the fact that it can be used for non-conducting samples [2]. The three modes used for AFM are contact mode, non-contact mode, and tapping mode [2]. In contact mode, the tip makes contact with the sample at all times. If the sample is a liquid, then surface tension forces brought about by adsorbed gases or the surface layer of the liquid can pull the tip into the sample. As a result, the tip can potentially damage the sample and dampen the tip. This mode is not the best setting for biological or protein samples. In tapping mode, the tip briefly makes contact with the surface and is immediately retracted. The typical oscillation is on the order of hundreds of kHz. The least popular mode of AFM is the non-contact mode where the tip is constantly a few nanometers away from the surface of the sample. Not surprisingly, the van der Waals forces detected on the tip and cantilever are much smaller [2].

In AFM, the data provided is a force curve, which is a plot of the force applied onto the tip versus the displacement of the tip. The location of the tip and movement of the AFM cantilever can be detected down to the nanometer thanks to the reproducible piezoelectric detection units. The movement of the cantilever can be used to compute the force applied using Hooke's law [18]. There are typically two plots provided by a force curve. When the tip is coming toward the surface of the sample, this is the approaching curve. When the tip is moving away from the surface of the sample, this is the retracting curve. With respect to noise, the two curves are overlaid on top of each other if there is no sample present. There should be two regions to these curves. There should be a horizontal region called the noncontact region that represents the baseline when a tip is not touching a surface. There also should be a region of the curve which rises up from the noncontact region called the vertical contact region which accounts for the van der Waals forces measured by the tip. The contact point is the spot on the force curve where the two regions meet and it is a critical portion of the force curve because it becomes a reference as to where the tip made contact with the surface [19].

The "polyprotein strategy" has been used to investigate proteins using AFM, especially the globular proteins [19]. This technique

involves placing our protein of interest into something called a multimodular construct. The modules within this construct have well recognized mechanical properties. In addition, they provide reference points for the location where the tip picks up the construct during each pulling event for the protein, and they function as linkers for pulling on the protein. Because globular proteins have dimensions on the order of nanometers, there are many a specific interactions between the surface and the tip. There is also no control on where the protein makes contact with the tip considering that most proteins are approximately 10 times smaller than the curvature of the tip. Thus, we are oblivious to the segment of the protein that is being stretched along with the direction that the force is being applied [19].

Another one of the more modern techniques used to study proteins is single-molecule force spectroscopy (SMFS). This technique was initially used to determine the binding energy and strength between a ligand and a receptor [18]. The AFM tip was modified to bind the receptor specifically while holding the ligand down on a support [18]. Moving the AFM tip close to the support created the ligand-receptor bond while moving it away broke the bond, known as the rupture force [18]. The force curve was used to measure the strength of this bond via determining the rupture force. Fundamentally, this technique is used to characterize surfaces. For protein applications, this technique can be used to provide knowledge about the conformation that a given protein may exhibit. SMFS is best suited to investigate intrinsically disordered proteins [19]. It has the benefit of exploring conformations that have millisecond lifetimes. Lastly, SMFS is capable of finding atypical conformations (under 10% of population) without adding special reagents to select for these specific conformations [19].

Sensitivity is one primary issue when performing SMFS techniques. It can distinguish a secondary structure within a protein; however, it cannot distinguish single hydrogen bonds within that secondary structure because it does not have the requisite force resolution. Thermal fluctuation of the cantilever due to Johnson noise is the primary limitation to force resolution. Johnson noise is the agitation of electrons caused by heat from the carriers of charge within an analytical instrument [2]. An improvement in the signal-to-noise ratio can be provided by reducing the size of the force sensor while maintaining the same rigidity. It is noted in the literature that the smaller cantilevers enable force resolution of approximately 7 to 10 pN, which provides a 2- to 3-fold improvement as compared to the standard size cantilevers. Nevertheless, these smaller cantilevers are not used because they are not widely available, not user-friendly and very expensive. Lock-in force spectroscopy is another approach to improving the force resolution. Schlierf et al. proposed this technique and it involves applying a small oscillation to the tip (about 5 nm amplitude) [20]. Detecting the magnitude of the oscillation moving through the polypeptide to the cantilever provides insight into the sample's elasticity. A force curve with lower noise and a resolution of 0.4 pN is obtained by using the elasticity signal and multiplying it by a reference signal [19].

Another issue with the SMFS techniques is the lack of throughput. A complete investigation of conformational equilibria for one set of experimental conditions can take multiple weeks. To achieve high-throughput assays, the experimental process will need to become automated and non-supervised. Next, data collection will need to be automated such that the few force curves that are significant can be pulled out from the thousands that are generated. The best SMFS instrument that has been used for high-throughput analysis has been published by Struckmeier et al. because the process for screening molecular interactions within the protein is streamlined [18].

Conclusion

In conclusion, an overview of some of the more modern techniques of analyzing proteins and peptides is provided. Bottom-up proteomics is becoming more commonplace to sequence proteins although it does not perform well with post-translational modifications. LC-MS is being used to separate compounds of interest before analyzing them with the MS detector to maximize the sensitivity of the compound. Ionization techniques are also used to clean up the sample matrix and labeling techniques are used to assist in quantifying these compounds. MALDI-IMS is an exciting new technique to help visualize the changes in the biology of a given tissue. Finally, AFM and SMFS were explored to help provide insight into the conformation changes within proteins. It is apparent that the techniques used to analyze proteins are diverse; however, they are now becoming more commonplace in current analytical research. Understanding how these techniques work will enable us to decipher the structure and function of proteins in a more efficient manner.

References

1. Webhoffer C, Schrader M (2011) Bioinformatic Tools for the LC-MS/MS Analysis of Proteins and Peptides: Protein and Peptide Analysis by LC-MS: Experimental Strategies, The Royal Society of Chemistry, Cambridge.
2. Skoog DA, Holler FJ, Crouch SR (2007) Principles of Instrumental Analysis. (6thedn), Brooks/Cole, Belmont.
3. Ackermann BL, Berna MJ, Eckstein JA, Ott LW, Chaudhary AK (2008) Current applications of liquid chromatography/mass spectrometry in pharmaceutical discovery after a decade of innovation. Annu Rev Anal Chem (Palo Alto Calif) 1: 357-396.
4. Bailey AO, Miller TM, Dong MQ, Vande Velde C, Cleveland DW, et al. (2007) RCADiA: simple automation platform for comparative multidimensional protein identification technology. Anal Chem 79: 6410-6418.
5. Egas DA, Wirth MJ (2008) Fundamentals of protein separations: 50 years of nanotechnology, and growing. Annu Rev Anal Chem (Palo Alto Calif) 1: 833-855.
6. Poole CF (2003) The Essence of Chromatography. Elsevier Science B. V., Amsterdam.
7. Thekkudan DF, Rutan SC, Carr PW (2010) A study of the precision and accuracy of peak quantification in comprehensive two-dimensional liquid chromatography in time. J Chromatogr A 1217: 4313-4327.
8. Moore AW Jr, Jorgenson JW (1995) Comprehensive three-dimensional separation of peptides using size exclusion chromatography/reversed phase liquid chromatography/optically gated capillary zone electrophoresis. Anal Chem 67: 3456-3463.
9. Hu S, Michels DA, Fazal MA, Ratisoontorn C, Cunningham ML, et al. (2004) Capillary sieving electrophoresis/micellar electrokinetic capillary chromatography for two-dimensional protein fingerprinting of single mammalian cells. Anal Chem 76: 4044-4049.
10. Vékey K (2001) Mass spectrometry and mass-selective detection in chromatography. J Chromatogr A 921: 227-236.
11. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, et al. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3: 1154-1169.
12. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, et al. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17: 994-999.
13. Ong SE, Mann M (2007) Stable isotope labeling by amino acids in cell culture for quantitative proteomics. Methods Mol Biol 359: 37-52.
14. Higgs RE, Knierman MD, Freeman AB, Gelbert LM, Patil ST, et al. (2007) Estimating the statistical significance of peptide identifications from shotgun proteomics experiments. J Proteome Res 6: 1758-1767.

15. Wiener MC, Sachs JR, Deyanova EG, Yates NA (2004) Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. Anal Chem 76: 6085-6096.

16. Burnum KE, Frappier SL, Caprioli RM (2008) Matrix-assisted laser desorption/ionization imaging mass spectrometry for the investigation of proteins and peptides. Annu Rev Anal Chem (Palo Alto Calif) 1: 689-705.

17. Otto M (1999) Chemometrics: Statistics and Computer Application in Analytical Chemistry. WILEY-VCH, Weinheim.

18. Struckmeier J, Wahl R, Leuschner M, Nunes J, Janovjak H, et al. (2008) Fully automated single-molecule force spectroscopy for screening applications. Nanotechnology 19: 384020.

19. Sandal M, Brucale M, Samori B (2010) Monitoring the Conformational Equilibria of Monomeric Intrinsically Disordered Proteins by Single-Molecule Force Spectroscopy: Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation. John Wiley & Sons, Inc., Hoboken.

20. Schlierf M, Berkemeier F, Rief M (2007) Direct observation of active protein folding using lock-in force spectroscopy. Biophys J 93: 3989-3998.