2
Literature Review

Proteins are very important biomacromolecules which are directly or indirectly involved almost all metabolic processes of living systems. Proteins are vital to many life processes such as cellular and structural repair, defense mechanism, metabolism of hormones and enzymes, molecular transportation, conducting stimuli, building nutrients, and energy production (Berg, Tymoczko et al. 2002). In addition to their critical role in cellular functions, proteins also play a very important role in maintaining the structural integrity of many tissues in the living systems (Berg, Tymoczko et al. 2002).

2.1 PROTEINS AND PROTEIN FOLDING

Fundamentally, proteins are the polymers of amino acids which are linked with each other through peptide bonds. There are around 20 different standard amino acids which make up almost all the proteins. During protein synthesis, the polypeptide chain of proteins, after its synthesis from ribosomal body, undergoes a protein folding pathway to attain its unique native state. Proteins are biologically active only when they attain their native 3-dimensional states (Nelson, Lehninger et al. 2008). However, sometimes proteins get misfolded which may trigger aggregation of soluble protein species (Chiti and Dobson 2009) (Figure 2.1). Cells possess certain molecular mechanisms that either repair or degrade misfolded proteins preventing aggregation. These molecular machineries are known as molecular chaperons, for example, Hsp70, Hsp40 and clustrins (Humphreys, Carver et al. 1999; Frydman 2001; Hartl and Hayer-Hartl 2002). The process of protein aggregation is very complex and mechanistic understanding of protein aggregation process is largely unknown. How does the process of protein aggregation begin? What are the changes in their conformational properties that trigger aggregation? Is there any way to prevent aggregation process? Clear answers to these questions are very much needed not only to understand the mechanism of protein aggregation but also to find some ways to prevent aggregation linked diseases.

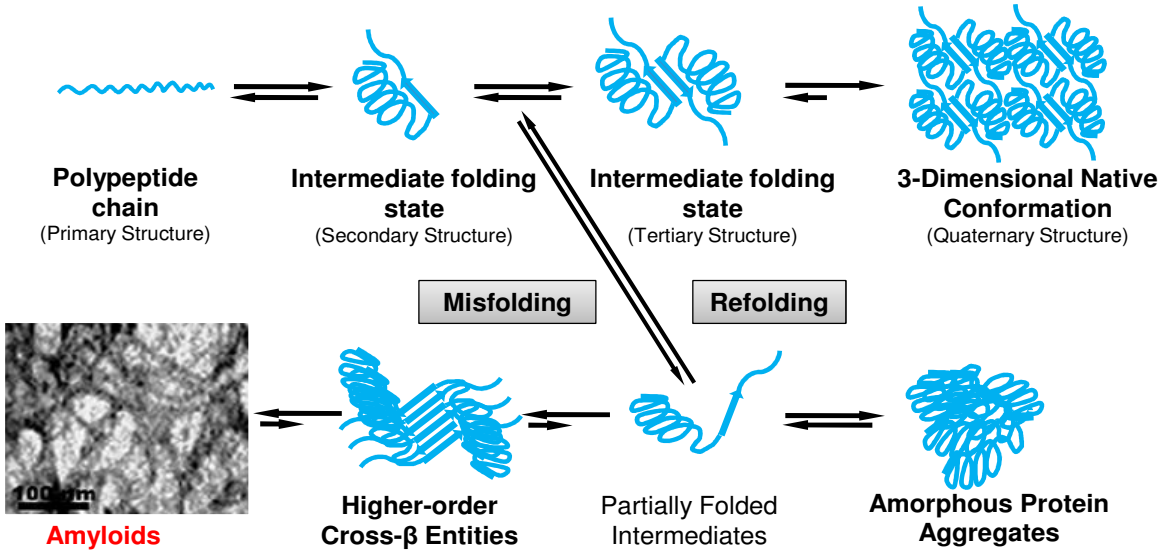

Figure 2.1 : Schematic representation of protein folding and misfolding process.

2.2 BIOLOGICAL SIGNIFICANCE OF SELF-ASSEMBLY OF COLLAGEN

Nature has employed the self assembly process of different proteins to generate higher order supramolecular structures which are very important to both structural and functional properties of living systems. Self-association of structural proteins such as collagen, keratin and actins, to form higher order assemblies from their respective soluble monomeric units is a fundamental process in biology (Berg, Tymoczko et al. 2002); (Nelson, Lehninger et al. 2008). Collagen is a structural protein which is known to have a unique triple-helical conformation. The molecular structure of collagen triple helix consists of three polyproline II like chains intertwined with each other through interchain H-bonds (Ramachandran and Kartha 1955; Rich and Crick 1961; Brodsky and Persikov 2005). The amino acid sequence of collagen has a unique sequence with frequent occurrence of $(Gly\text{-}X\text{-}Y)_n$ repeating patterns. One of the characteristic features of collagen triple helical molecules is to undergo a self-assembly process and to form unique higher order supramolecular structures such as fibers, tendons and networks (Kielty and Grant 2003). It is believed that the driving force for the self assembly process of triple helical molecule involves non-covalent interactions including hydration and hydrophobic interactions (Leikin, Rau et al. 1995; Leikin, Parsegian et al. 1997; Kar, Amin et al. 2006). As of now, ~28 distinct types of collagens have been identified. However, functional properties are known for some. Among different types of collagens, type I is one of the most abundant form and it is found in tissues as regular D-periodic fibrils (Hulmes, Miller et al. 1973; Hofmann, Fietzek et al. 1980; Brodsky and Persikov 2005). In addition to fibrillar collagens many non-fibrillar forms of collagens are present which make up vital tissue components in the body system. Such non-fibrillar collagens include networks of type IV collagen in the basement membrane (Yurchenco and Ruben 1987; Yurchenco and Ruben 1988; Yurchenco and Schittny 1990), hexagonal networks of type VIII collagen in the subendothelial layers (Stephan, Sherratt et al. 2004), antiparallel arrays of type VII collagen (Burgeson 1993) and microfibrils of type VI collagen in all connective tissues (Engel, Furthmayr et al. 1985). Though collagens are vital for proper functioning of tissues but sometimes they are also known to cause several pathologies. For example, excess platelet aggregation mediated by fibrillar collagens can cause thrombosis which leads to myocardial infarction and stroke (Farndale, Sixma et al. 2004). Type VIII collagen is involved in atherosclerosis lesions and plaque formation (Plenz, Deng et al. 2003). When these plaques rupture they form platelet thrombus which causes sudden heart attack if thrombus is formed in the coronary arteries or it may cause stroke if thrombus formation is in cerebral artery (Plenz, Deng et al. 2003). Blockage of cardiac arteries due to collagen accumulation at the site of angioplasty procedure is known to cause restenosis (Lafont, Durand et al. 1999). Since excess collagen accumulation is directly linked to severe pathologies, it is important to find inhibitors against collagen fibril formation. It has been already reported that sugars and polyols are capable to inhibit the fibrillogenesis of type I collagen by disrupting hydrogen-bonded water bridges between the helices (Kuznetsova, Chi et al. 1998). It is also known that sugars are able to inhibit self-association of collagen triple-helical model peptides (Kar, Amin et al. 2006). Considering all these factors, it is important to find potential molecules that can interfere with the intermolecular interactions between collagen molecules and to inhibit collagen fibril formation. Identification of these inhibitors may help in the development of therapeutics against collagen activated diseases.

2.3 AMYLOID FORMATION AND ITS RELATED PATHOLOGIES

The process of amyloid formation of certain proteins involves conversion of soluble proteins into insoluble higher order cross-β entities. Amyloid formation of proteins is sometimes known to be beneficial and essential for certain biological functions such as, pigmentation of skin and eye, cell-cell contacts in *Escherichia coli*, and storage granules in pituitary gland (Chapman, Robinson et al. 2002; Berson, Theos et al. 2003; Maji, Perrin et al.

2009) . However, on the other hand, the process of amyloid formation has been mostly implicated in several pathological complications including many neurodegenerative diseases (Chiti and Dobson 2009).

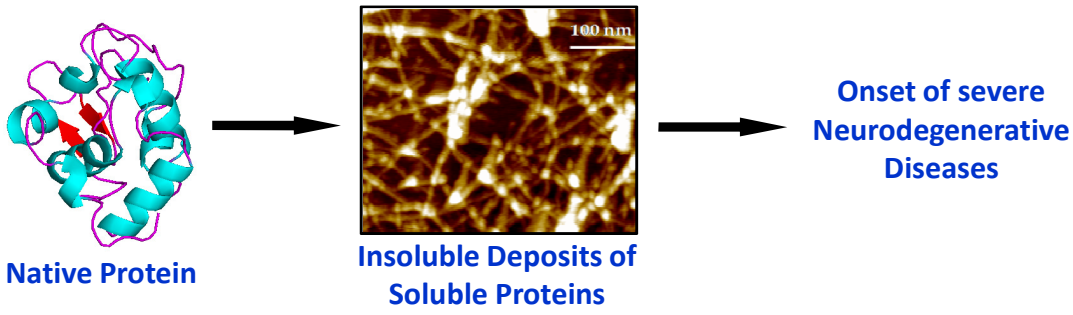

Figure 2.2 : Conversion of soluble native protein into insoluble amyloid structures is linked to many neurodegenerative diseases.

Until now, around 40 different proteins have been known to be associated with amyloid-linked diseases. Protein amyloids are known to form cytotoxic entities such as formation of plaques and tangles in Alzheimer's disease, lewy bodies in Parkinson's disease and intranuclear inclusions in Huntington's disease (Soto 2003).

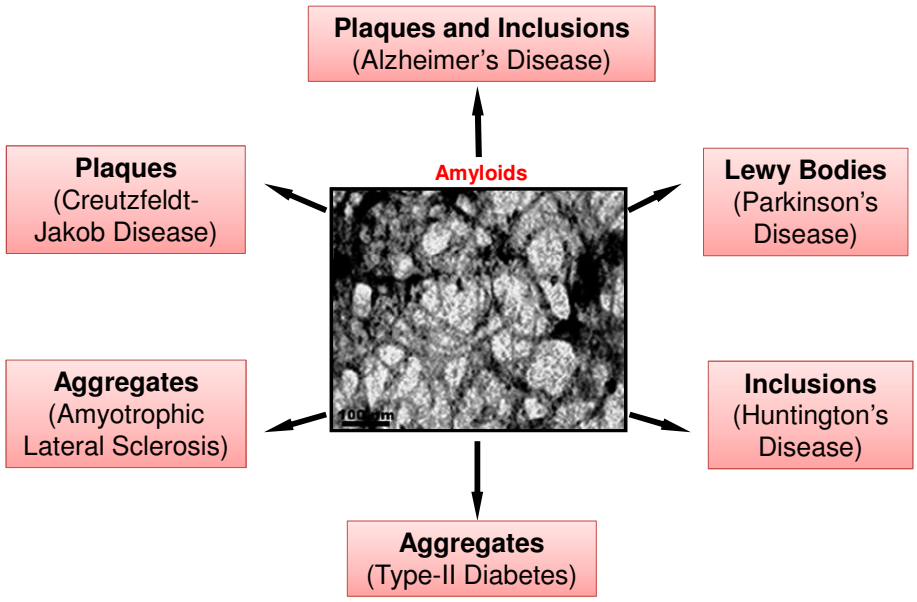

Figure 2.3 : Amyloid formation leads to severe lethal diseases.

To understand the disease mechanism associated with protein aggregation it is important to understand the nucleation event that begins protein aggregation. For many proteins, the process of amyloid formation is believed to follow a nucleation dependent pathway (Lomakin, Chung et al. 1996; Xue, Homans et al. 2008). Monomers may interact to form stable oligomers, and if critical concentration of oligomer population is achieved it may trigger the aggregation process (Lomakin, Chung et al. 1996; Xue, Homans et al. 2008). In some cases, the protein monomer itself can act as a nucleus such as in the case of aggregation of polyglutamine peptides into amyloid fibrils (Wetzel 2006). The aggregation profile of any protein which follows a nucleation growth mechanism is believed to comprise of three distinct phases: **lag phase**, time taken by proteins monomers to form oligomers; **growth phase**,

exponential growth of the formation of aggregates; and **plateau phase**, indicating completion of aggregation (Figure 2.4).

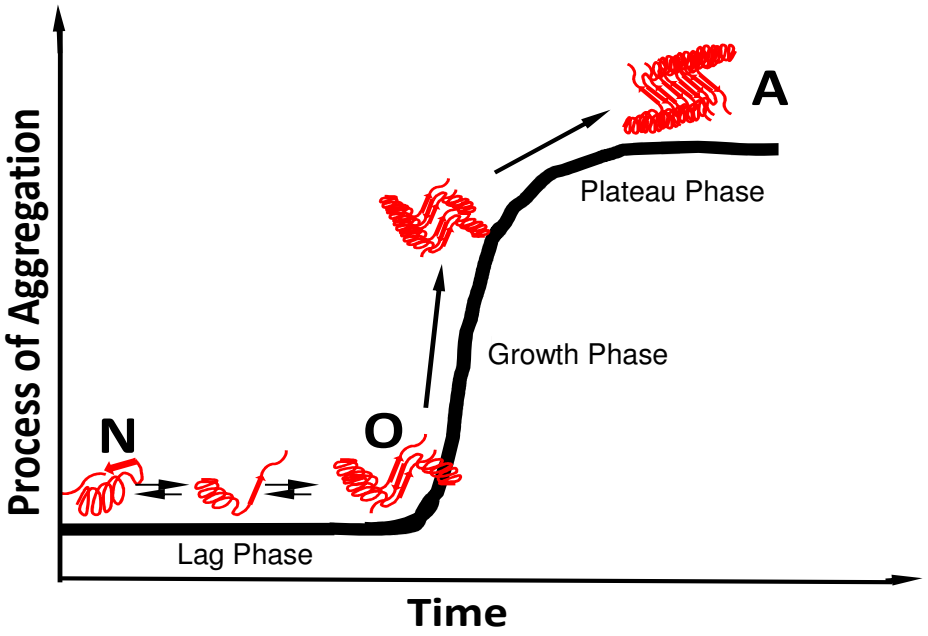

Figure 2.4 : Schematic representation of nucleation process of protein aggregation as a function of time. Where, N, is the native conformation of a protein; O, is the oligomeric state of it; and A, is the aggregated mature higher order structures.

Recent studies reveal that the on-pathway oligomers are more toxic than the mature amyloid fibers. Therefore in order to understand the process of protein aggregation researchers are now more interested in understanding the process of oligomerisation and their cytotoxic effects. It has been also reported that the structural conformation of oligomers are typical cross-β type (Apetri, Maiti et al. 2006; Holm, Jespersen et al. 2007; Chiti and Dobson 2009). When proteins undergo an aggregation process, they generally form highly stable higher order structures with diverse morphologies such as, disordered amorphous aggregates and highly ordered cross-β structured amyloid fibrils (Tyedmers, Mogk et al. 2010) (Figure 2.1). The cross-β conformation is usually achieved by non-covalent interactions between two partially unfolded chains. The partially unfolded protein species are known to trigger self-association of proteins. Such interactions are believed to be facilitated by hydrophobic-hydrophobic interactions mediated through the exposed hydrophobic groups (Schmittschmitt and Scholtz 2003). The cross-beta structure is usually stabilized by stable H-bonds mediated through the functional groups of the backbone of the polypeptide chain. Further these cross-beta structures are also stabilized by interactions between the functional groups of the side chains which includes pi-pi stacking, hydrophobic interactions and salt bridges (Gazit 2002; Tartaglia, Cavalli et al. 2004). All these non-covalent interactions are vital to the formation of supramolecular protein assemblies.

Till now it is reported that the aggregation of particular protein is associated with a specific clinical pathology such as, Aβ amyloid formation leads to plaque formation in Alzheimer's disease. Though many proteins are known to form amyloid fibers, the cross-beta structure becomes a common structural feature for all these fibers (Tyedmers, Mogk et al. 2010).

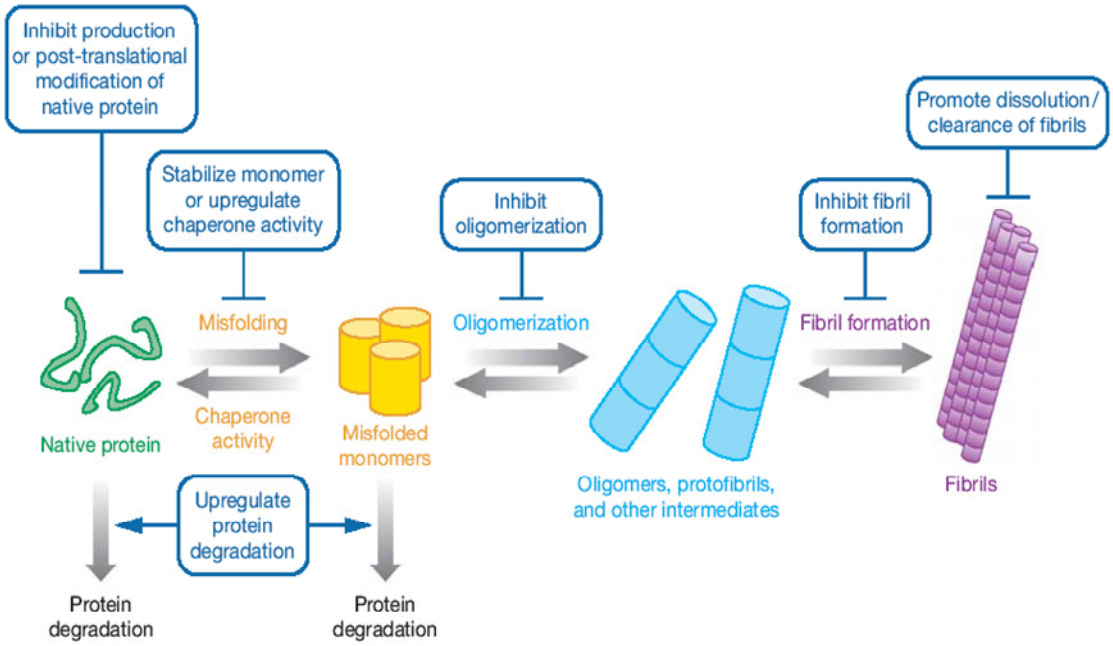

Figure 2.5 : Therapeutic strategies against amyloidogenic diseases. (Skovronsky, Lee et al. 2006)

2.4 STRATEGIES TO TARGET PROTEIN AGGREGATION PROCESS

Usually globular proteins such as, lysozyme, serum albumin, and insulin, are considered as convenient model systems to study protein aggregation in *in vitro* system. These globular proteins are well studied and much information is available in literature that covers both structural and functional properties (Swaminathan, Ravi et al. 2011; Gong, He et al. 2014; Li, Wu et al. 2016). Aggregation of lysozyme may cause systemic amyloidosis (Pepys, Hawkins et al. 1993) whereas insulin aggregation is known to trigger site-specific localized amyloidosis (Dische, Wernstedt et al. 1988). Amyloid formation of insulin is also a great concern for its storage as therapeutics agents (Hjorth, Norrman et al. 2016). These approaches involve: i) inhibition of oligomerization, ii) stabilization of protein monomers, and iii) destabilizing the formed oligomers and mature fibrils. Hence, globular proteins can be used as convenient model system to study amyloid formation and finding inhibitor against such process is important for amyloid research. There are many possible approaches by which one could target amyloid formation, as illustrated in the Figure 2.5.

Using a combination of different biophysical techniques one could gather valuable insights which may help us to understand the process of protein aggregation. Therefore, to carry out *in vitro* amyloid formation certain biophysical techniques seems to be very beneficial. One of the fundamental experiments to monitor amyloid formation in proteins is the Thioflavin T assay (ThT) (Krebs, Bromley et al. 2005). ThT has strong affinity to bind amyloid fibrils and such binding is known to enhance its fluorescence property (Krebs, Bromley et al. 2005). Therefore, formation of amyloid fibrils of a protein can be studied by monitoring the rise in ThT signals (Krebs, Bromley et al. 2005). Circular Dichroism helps to understand the secondary structure contents of proteins and also to understand their thermal stability (Whitmore and Wallace 2008). FTIR spectroscopy is also an important tool to study the secondary structure of the proteins both in their molecular and fibril forms (Kong and Yu 2007). Apart from all these techniques, dynamic light scattering (DLS) technique usually employed to analyze the particle size of the proteins (Zhang and Smith 1993). Morphologies of different higher order assemblies formed during protein aggregation can be visualized using different microscopic tools such as, atomic force microscopy (AFM), transmission electron microscopy (TEM) and scanning electron

microscopy (SEM) (Alessandrini and Facci 2005). In addition to these biophysical techniques *in silico* experiments including molecular docking studies are very helpful to understand protein-protein interactions as well as protein aggregation (Irwin and Shoichet 2005; Brooks, Brooks et al. 2009). The information obtained from both experimental and computational studies on protein aggregation may improve our foundational knowledge which may perhaps inspire researchers working on cell models or animal models. Therefore, studies involving amyloid formation of proteins in *in vitro* system become vital to the mechanistic understanding of the aggregation process and to the development of therapeutics against amyloid-linked diseases.

...