2
Literature Review

Proteins are the most important macromolecules in nature which are directly or indirectly involved in almost every metabolic process in living systems. These proteins are made up of amino acids connected to each other via peptide bonds. Once the polypeptide chain is synthesized from ribosome, the nascent polypeptide chain undergoes a process of protein folding to arrive at the destined three-dimensional conformation which is critical to the functional properties of the protein. Proteins act as building blocks and regulate all the cellular functions as well as structural integrity of an organism. Humans produce a range of approximately 30,000 different proteins, for specific roles (Herczenik and Gebbink 2008). These proteins play various roles in the organism such as cellular and structural repair, defense mechanism, metabolism of hormones and enzymes, molecular transportation, conducting stimuli, building nutrients and energy production in a cell (Berg JM, Tymoczko JL 2002).

2.1 PROTEIN AGGREATION AND ITS BIOLOGICAL SIGNIFICANCE

Though proteins fold into their native soluble structures, in some cases due to many factors they tend to self-assemble and form higher order structures. The conversion of protein molecules from their soluble state into well-organized higher order structures such as collagen fibrils and amyloid entities is a fundamental process in biology. Aggregation of soluble proteins into amyloid fibrils which display the existence of cross β structures, is known to possess distinct biophysical and histological properties. Further the process of formation of protein amyloids is believed to cause several pathologies including a series of neurodegenerative diseases. However, in some cases, amyloid formation of proteins is also known to be beneficial and essential for certain biological functions. Such functional amyloids are predominant in pigmentation of skin and eye, cell-cell contacts in *Escherichia coli* and storage granules in pituitary gland (Chapman et al. 2002)(Berson et al. 2003)(Maji et al. 2009). Other than biological significance, these amyloids are also used in construction of nanomaterials like nanowires (Reches 2003), hydrogels (Yan et al. 2008) and biocompatible scaffolds for cell growth (Jacob et al. 2015)(Das et al. 2016).

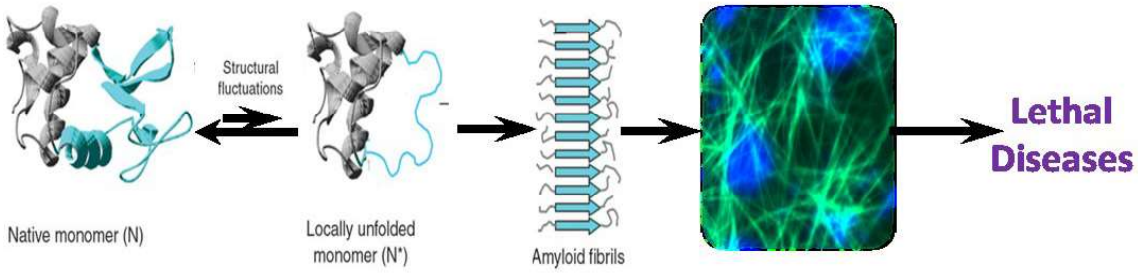

Figure 2.1: Conversion of soluble native protein into insoluble well organized amyloid structures linked to various diseases adapted from Chiti and Dobson 2009.

Though protein aggregation process is a fundamental process in Biology (W. Wang 2005), in many cases aggregation of proteins into amyloid fibrils becomes one of the foundational causes for the onset of complicated diseases. The process of amyloid formation has been mostly implicated in neurodegenerative diseases (Dobson 2003)(Chiti and Dobson 2009). Till today approximately 40 different proteins are known to form amyloid like entities. Formation of plaques and tangles in Alzheimer's disease, Lewy bodies in Parkinson's

disease, and intra-nuclear inclusion bodies in Huntington's disease have been known to be linked to β-amyloid aggregates of the related proteins or peptides (Soto 2003) figure2.2. Recently it was discovered that single metabolites such as aromatic amino acids can also self-associate and form fibrils that have amyloid-like biophysical, biochemical and cytotoxic properties (Adler-Abramovich et al. 2012) (Singh et al. 2014)(Shaham-Niv et al. 2015).

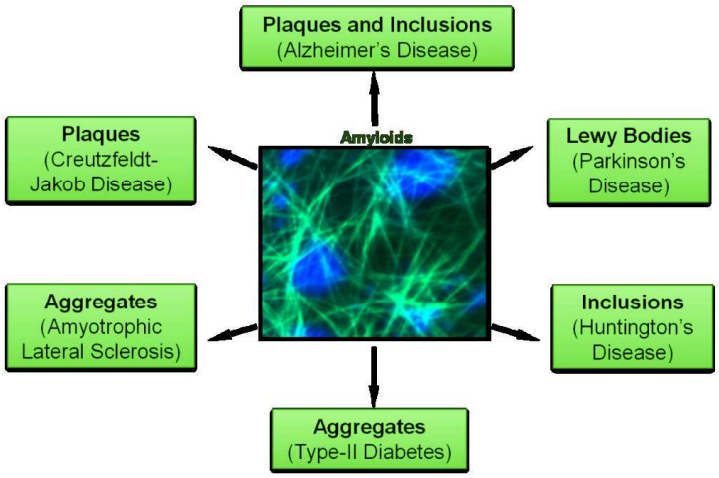

Figure 2.2: Schematic representation of various pathologies linked to formation of amyloid fibrils

The process of protein aggregation is also a major problem during DNA-recombinant synthesis (Dunker et al. 2002). Additionally, protein aggregation is a serious problem during storage of protein therapeutic agents, such as insulin (Liza Nielsen et al. 2001). Due to all these factors, it is important to have a mechanistic understanding of the protein aggregation process. It is well known that the nucleation event triggers the protein aggregation process. Protein misfolding is a frequent intrinsic structural alternation within the protein molecule and such misfolding process normally is accelerated by mutations (change in amino acid compositions) and other environmental factors like variation of temperature, pH and hyperglycemic conditions (Herczenik and Gebbink 2008). Under such extreme conditions proteins can lose their native conformation more rapidly. These unfolded or partially unfolded protein species are known to have an inherent tendency to aggregate (Figure 2.1), as explained by Chiti and Dobson (Chiti and Dobson 2009).

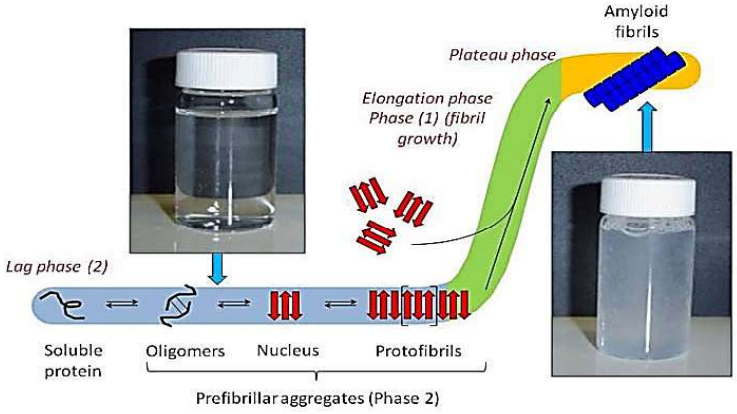

Figure 2.3: Schematic representation of the mechanism of protein aggregation under *in vitro* conditions.
CREDIT*: Butterfield, S. M. & Lashuel, H. A. Angew. Chemie (2010)*

In most cases, the protein aggregation is considered as a nucleation dependent process (Lomakin et al. 1996)(Xue et al. 2008). It has been proposed that under aggregating conditions, the population of aggregation prone partially folded species play a vital role for the onset of aggregation (Uversky and Fink 2004). Native protein species are also known to undergo aggregation when present at a very high concentration (Nelson et al. 2005). At critical concentrations, sufficient to trigger aggregation, the process of aggregation usually follows distinct lag phase, log phase and a stationary phase (Butterfield and Lashuel 2010) (Figure 2.3). In the lag phase the soluble monomers are converted to oligomers and these oligomers may act as nucleating agents in which the protein monomers get reversibly attached to the growing core and form protofibrils. In the log phase, the protofibrils further self-assemble and form well organized linear branched or unbranched fibrils. The final phase is the stationary phase that indicates the completion of the growth of amyloid fibrils. Under *in vitro* conditions, the process of amyloid aggregation can be studied by using amyloid specific dyes: Congo red and Thioflavin T (Wilcock et al. 2006). Thioflavin T is a fluorescent dye that is known to bind to the cross-β structured amyloid fibrils (Groenning 2010) and in its amyloid-bound form its fluorescence property increases. Hence by monitoring the increase in the Thioflavin T fluorescence signal one could obtain the kinetics of any protein undergoing amyloid aggregation. The cross-β conformation is usually achieved by non-covalent interactions between two partially unfolded chains. Such interactions are believed to be facilitated by hydrophobic-hydrophobic interactions mediated through exposed hydrophobic groups (Schmittschmitt and Scholtz 2003). The cross-beta structures are usually stabilized by stable H-bonds mediated through functional groups to the backbone of 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 et al. 2004). All these non-covalent interactions are vital to the formation of supramolecular protein assemblies

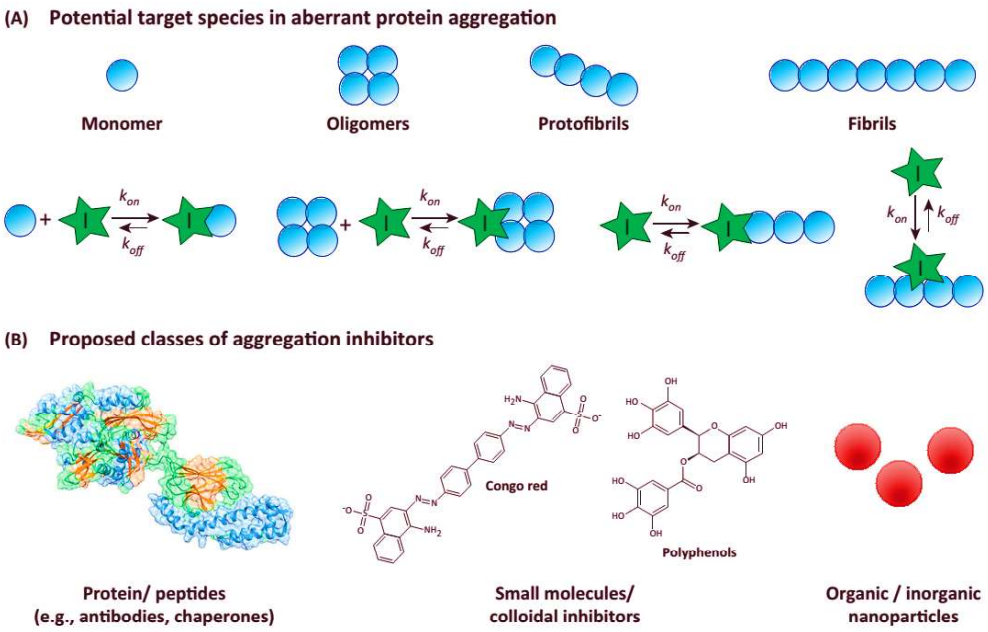

Figure 2.4 *Strategies to target Protein aggregation.* Credit: Paolo Arosio, et.al Cell press 2014.,

Usually globular proteins such as lysozyme, serum albumin and insulin are considered as convenient model systems to study protein aggregation under *in vitro* conditions (Dubey et al. 2014). These globular proteins are well studied proteins and much information is available in

literature that covers both structural and functional properties (Gong et al. 2014)(R. Li et al. 2016)(Swaminathan et al. 2011). Aggregation of lysozyme has been linked to systemic amyloidosis (Pepys et al. 1993) whereas insulin aggregation is known to trigger site-specific localized amyloidosis (Dische et al. 1988). Amyloid formation of insulin is also a great concern for its storage as therapeutics agents (Hjorth et al. 2016). Hence, globular proteins can be used as convenient model systems to study amyloid formation and finding inhibitor against such process is important for amyloid research. Prevention of aggregation process can be achieved by many possible approaches as illustrated in the Figure 2.4. These approaches involve: i) inhibition of oligomerization, ii) stabilization of protein monomers and iii) destabilizing the formed oligomers and mature fibrils.

2.2 EFFECTIVE STRATEGIES TO TARGET PROTEIN AGGREGATION PROCESS

One of the most effective strategies for targeting amyloid-linked diseases is to find potential inhibitors against the initiation of the process of amyloid formation. Successful designing of effective inhibitors of protein aggregation process has gained much attention in recent years. Several biomolecules including single molecules, amino acids, natural compounds, peptides, proteins and nanoparticles have been reported to act as inhibitors of amyloid fibril formation of different proteins (Ghosh et al. 2009)(Kar and Kishore n.d.)(Shiraki et al. 2002)(Rajasekhar and Govindaraju 2015) (Viet et al. 2011)(Zaman et al. 2014). Besides, it has been observed that the nanoparticles are more efficient to stimulate, to inhibit, or to delay the fibrillation kinetics. In recent years, there is an increasing interest in the making of novel nanoparticles that are now extensively utilized in Chemistry, Biology, Engineering, and Medicine. These efficient nanomaterials are expected to bring about a revolution for the inhibition of the aggregation process of proteins and peptides (Giljohann et al. 2010). Over the past decade much research has also focused on the surface functionalization of metallic nanoparticles with potential compounds to target amyloid formation of proteins (Álvarez et al. 2013)(Siposova et al. 2012)(Palmal, Maity, et al. 2014). Some studies have suggested that nanoparticles coated with hydrophobic molecules can inhibit the fibrillation process of proteins (Dubey et al. 2015). Hydrophobic nanoparticles or nanoparticles functionalized with hydrophobic molecules such as KLVFF peptide (Richman et al. 2011), phe-phe dipeptide (Skaat et al. 2012), and di hydrolipoic acid (Thakur et al. 2011) have been reported to inhibit the fibrillation process of different proteins. Further, nanoparticles functionalized with specific molecules such as curcumin, dextran, and sialic acid are also known to inhibit protein aggregation (Kouyoumdjian et al. 2013)(Mahmoudi et al. 2013)(Mathew et al. 2012)(Palmal, Maity, et al. 2014)(Taylor et al. 2011). These results suggested that further detailed studies are required for better understanding of the effects of modular surface chemistry on the process of amyloid fibril formation. Various properties of the nanoparticles, such as their size and thermal stability, are critical factors for their inhibition effect against amyloid fibril formation of proteins (Zaman et al. 2014). Some studies have also reported that the anti-amyloid activity of inhibitors is greatly enhanced when these inhibitor molecules are surface functionalized with the nanoparticles(Anand et al. 2016)(Dubey et al. 2015)(Palmal, Jana, et al. 2014).

2.3 COLLAGEN SELF-ASSEMBLY AND ITS BIOLOGICAL SIGNIFICANCE

In addition to amyloidogenic proteins, there are many structural proteins such as collagens in the body system that undergoes self-assembly process to arrive at important higher order structures. Collagen is the most abundant protein present in mammals and contributes almost 25% of the total protein. The collagen molecule consists of three polyproline II chains which are intertwined with each other by interchain hydrogen bonds, yielding a right handed triple helical structure (Rich and Crick 1961); (Ramachandran and Kartha 1955); (Brodsky and Persikov 2005). The polypeptide chain of collagen molecule is composed of Gly-Xaa-Yaa repeats where X and Y positions are repeatedly occupied by proline and hydroxyproline residues

respectively. (Gly-Pro-Hyp)n repeating regions in collagen have been reported to contribute for both molecular stability and self-assembly process of collagen. In the tissues of their existence collagen triple helical molecules are known to undergo a self-assembly process and such process is vital to the structural and functional properties of many tissues. It has been reported that the non-covalent interactions including hydration and hydrophobic interaction act as the main driving force for the self-assembly process of collagen molecules (Kar et al. 2006) (Leikin et al. 1995)(Leikin et al. 1997) to yield fibers, tendons and networks (Kielty and Grant 2003). Till today approximately 28 different categories of collagens have been identified based on their place of existence, functions and their structural conformation. Among different types of collagens, type I is one of the most abundant forms and it is found in tissues as regular D-periodic fibrils (Hofmann et al. 1980)(David J.S. Hulmes et al. 1973). In addition to fibrillar collagens like type I, II, III, V, XI, XXIV, and XXVII (Fratzl 2008)(D.J.S. Hulmes 1973) 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) and hexagonal arrays of type VIII collagen in endothelial tissues.

Though collagens are vital for proper functioning of tissues, excess collagen fibril formation is known to cause several pathologies. For instance, excess platelet aggregation facilitated by fibrillar collagens can cause thrombosis which causes stroke and myocardial infarction (Farndale et al. 2004). Excess accumulation of type VIII collagen causes atherosclerosis lesions and plaque formation (Plenz et al. 2003). Rupture of plaques leads to the formation of thrombus inducing heart attack or stroke (Plenz et al. 2003). The deposition of excess collagen around the cardiovascular stents at the site of angioplasty causes restenosis (Lafont et al. 1999). Since excess collagen accumulation is linked to severe pathologies, it is important to design inhibitors against collagen fibril formation. It has been already reported that sugars and polyols are capable of inhibiting the fibril formation of type I collagen by disrupting hydrogen-bonded water bridges between the helices(N. Kuznetsova et al. 1998). It is also known that sugars can prevent the self-association of collagen triple-helical model peptides (Kar 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 suitable inhibitors may help in the development of therapeutics against collagen activated diseases.

2.4 BIOPHYSICAL TOOLS TO STUDY PROTEIN AGGREGATION

Using a combination of different biophysical techniques, one can 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 (Krebs, Bromley et al. 2005). Thioflavin T has strong affinity to bind amyloid fibrils and such binding is known to enhance its fluorescence property (Krebs, Bromley et al. 2005). Hence, formation of amyloid fibrils of a protein can be studied by monitoring the rise in Thioflavin T 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 forms as well as in their fibrillar 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 2008) that mostly helps in monitoring the formation of higher order structures during aggregation process. 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 et al. 2005). In addition to these biophysical techniques *in silico*

experiments including molecular docking studies are very helpful to understand protein-ligand interactions as well as protein aggregation (John J. Irwin et al. 2005)(Brooks et al. 2009).

Many established protocols can be used for making of nanoparticles, particularly for synthesis of surface-functionalized nanoparticles. Reducing agents such as tri-sodium citrate, ascorbate, sodium borohydride ($NaBH_4$), potassium hydroxide (KOH), Tollens reagent, N, N-dimethylformamide (DMF), and poly (ethylene glycol)-block copolymers are generally used for the synthesis of silver and gold (Iravani et al. 2014)(J. Kimling et al. 2006)(Herizchi et al. 2014) to yield a suspension of homogeneous nanoparticles. All the nanoparticles synthesized can be characterized by techniques such as UV visible spectroscopy, fourier transform infra-red spectroscopy, X ray diffraction spectroscopy, atomic force microscopy, scanning electron microscopy and transmission electron microscopy (Anand et al. 2015)(Šileikaitė et al. 2009). It is also necessary to test the biocompatibility of the synthesized nanoparticles. One of the effective strategy is to check their effect on hemolysis assay, which can reveal the hemocompatibility of the functionalized nanoparticles (Vignesh et al. 2015).

The information obtained from both experimental and computational studies on protein aggregation process in the presence of strategically designed nanoparticles may improve one's 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 nanoparticle based therapeutics against amyloid-linked diseases.

…