CHAPTER 1

INTRODUCTION AND REVIEW OF LITERATURE ON PROTEIN FOLDING, AMYLOIDOSIS AND NEURODEGENERATION*

Abstract

Protein folding in general and subsequent mechanisms that lead to its misfolding and further complications such as aggregates formation are still not fully understood. Over the years, protein misfolding and amyloidosis have been deciphered as one of the leading causes of many neurodegenerative diseases. Interestingly, various non-neuropathic conditions arise due to faulty protein folding mechanisms. Some of the prominent clinical conditions due to amyloidosis include Alzheimer's Disease (AD), Parkinson's Disease (PD), Huntington's Disease (HD), Kuru's Disease and Creutzfeldt-Jakob Disease, which are majorly neuropathic. Among non-neuropathic conditions, A-Light Chain Amyloidosis, Lysozyme Amyloidosis, and Fibrinogen Amyloidosis are very common. Currently, there are very few treatment regimens for all of these conditions. There are very few FDA-approved drugs for acute conditions, so there is an increasing mortality rate, especially among the geriatric population. This has led to a surge in finding novel therapeutic options, primarily through two strategies: drug repurposing and finding a natural cure by extensive stress on extracts from medicinal plants and or small molecules. In the current project, we have tried to connect the link between cellular homeostasis losses due to intracellular metabolites imbalance. Purines and Pyrimidines have long been questioned in their connection with such protein misfolding diseases as they are involved in critical biochemical reactions connected with proper proteostasis. We now include some *in silico* and *in vitro* studies to elucidate the role of individual key nitrogenous bases and nucleosides and their role in protein folding and aggregation.

---

* Part of the work is published in Kundu *et al*., (2020) Advances in protein misfolding, amyloidosis and its correlation with human diseases; 3 Biotech. 10(5):193.

1.1 Introduction to Protein Folding

One of the biochemists' fundamental study topics is the inherent property of polypeptide chains to fold to a native biologically functional state in a brief period (Luheshi and Dobson, 2009). One of the fundamental questions that biologists raise is how an unfolded polypeptide finds its way through to the native structure in a biologically reasonable time without possibly passing through all the conformations, also known as "Levinthal's paradox". Cyrus Levinthal, in 1969, in order to understand how proteins fold in such a short time, reaching global energy minima, presumed that the protein has to go through all the conformations, reach individual energy minima, and check for stability, finally gaining the native structure. Even if a globular protein has 100 amino acids, the number of conformations for all the 100 amino acids will be astronomically colossal: 2^{100}, implying that the number of conformations it has to pass through is 2^{100} to attain the biologically native structure. This was impossible even if a protein changes its conformation every 1ps, the fastest time at room temperature given by thermal vibrations (fastest physical process at room temperature). Even in this case, the time taken for all 2^{100} conformations is 10^{10} years approximately. This calculation gave birth to the paradox. Levinthal concluded that if a protein folds and reaches the native structure, which is also the global energy minima, as stated by Afinsen after his experiment, going through all the possible conformations, then the native structure of any protein may not necessarily be the global minima. Levinthal further stated that protein folding is preceded by some unique pathways identified as a part of the evolution and the native structure thus attained is just the end of that unique pathway (Levinthal, 1968, Privalov, 1979, Ivankov and Finkelstein, 2020). The folding funnel or the "energy landscape" theory states that any unfolded polypeptide is at the top of the funnel that descends to the funnel's centre as energy is further lowered by the subsequent formation of various intermolecular interactions. The number of possible conformations the protein must go through to reach the native state is also reduced, thus simplifying the overall process.

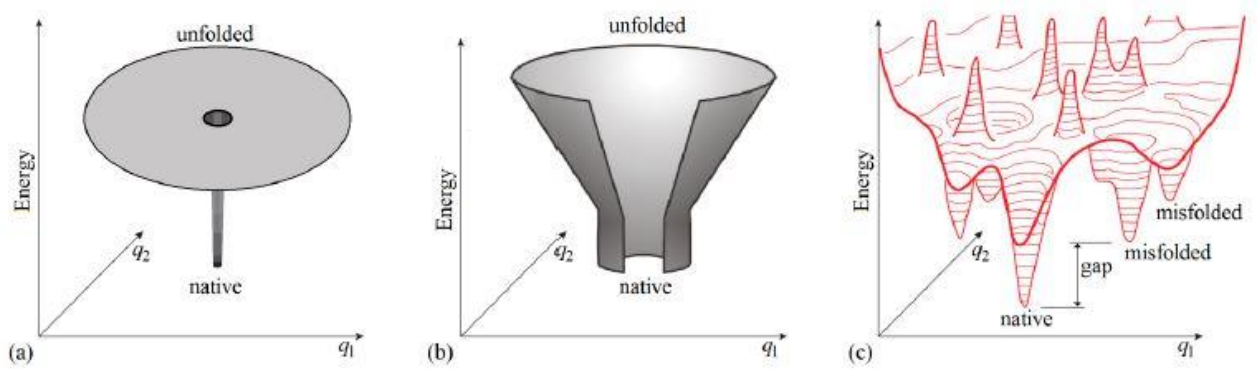

Fig 1.1 Different models to explain the protein folding process: For easy understanding, q1 and q2 represent the entire conformational space containing hundreds of dimensions. **(A)** The golf course model, the closest physical representation to Levinthal's explanations, represents the protein in a ball rolling over a flat golf course with random sample conformation, ultimately hitting the golf hole, representing the native structure. **(B)** Fig b represents the general energy funnel. The funnel's centre represents the global energy minima, and **(C)** Fig c represents a more detailed protein folding energy landscape. The "gap" denotes the energy difference between global and other energy minima necessary to cross to transition and unfolded chain structures into the native state. The energy gap is defined by k_BT (melt), where k_B represents the Boltzmann constant, and T(melt) is the melting temperature of the protein (Adapted with permission from Ivankov and Finkelstein, 2020).

Although in moderately large proteins with amino acids 150-200 in number, the funnel's folding process is not entirely smooth rather than rugged due to the number of intermediate states present. During protein folding, the protein may get trapped in such intermediate states after attaining local energy minima, and a specific energy "gap" must be crossed to reach the bottom of the funnel where the native structure is present. These transient (Figure 1.1) partially folded protein state conformations in the energy landscape exhibit good hydrophobic patches, often subjecting them to self-assembly (Oberg *et al.*, 1994).

1.2- History and clinical relevance of amyloidosis

The first evidence of amyloids in the pretext of medicine was in 1639, later reviewed by Cohen (Cohen *et al.* 1986). Rudolf Virchow conducted extensive studies examining wax-like deposits from organs like the liver, kidney, and spleen. He later followed up this study by naming such wax-like deposits as "amyloid" and introducing the word to the medical fraternity. He also observed extensive tissue abnormalities in cerebral corpora amylacea (Virchow *et al.* 1854, Kumar *et al.* 2009). Much later in the early 20th century, Alois Alzheimer, one of the finest German psychiatrists, worked on a patient Auguste Deter who presented with short-term memory loss symptoms (Figure 1.2). After the patient died in 1906, Alzheimer investigated her medical reports. Alzheimer and two Italian physicians used the Bielschowsky staining technique and identified

amyloid plaques (senile plaques) and neurofibrillary tangles. These brain structures were later identified as the leading cause of Alzheimer's Disease (AD). The case of Auguste Deter was the first reported incidence of AD (Forman *et al.* 2004). One of the other most commonly occurring neurodegenerative diseases, Parkinson's Disease (PD), was later investigated by Freidrich Lewy and reported similar pathological marks, named Lewy Bodies and Lewy Neuritis (Lewy *et al.* 1912).

There are various factors often designated to affect *in vivo* protein folding. Factors like the protein concentration, which is synthesized, and the molecular mechanisms available *in vivo* like molecular chaperones, proteases, and molecular crowding are some of the factors studied regularly in protein aggregation (Minton, 2001). Fact that the native structure of the protein is of utmost importance to make it biologically functional, any anomalies in the native structure ultimately lead to misfolded protein and awkward transitions in the native structure, enhancing the probability of forming protein aggregates (Goldberg *et al.*, 2003, Konar *et al.* 2019). Structural transition in the native protein results in hydrophobic amino acid residues' exposure to the solvent instead of the protein's core. This major transition forms the seed, escalating the formation of well-ordered, large, insoluble fibrillar aggregates. These aggregates are amyloid fibrils (Kundu *et al.*, 2020).

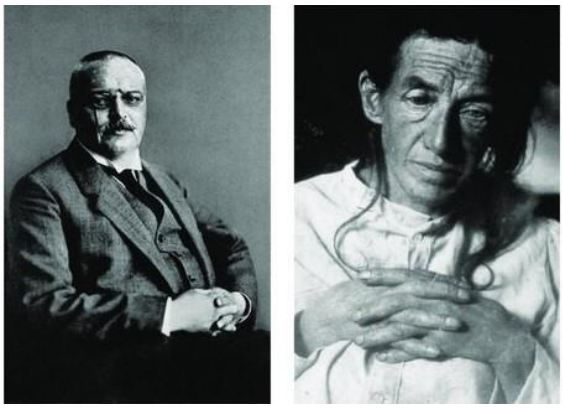

Fig 1.2 History of Alzheimer: Alois Alzheimer and his patient, Auguste Deter, the first person reported to have died from Alzheimer's disease (AD) [Source: Wikipedia]

1.3- Intermediate Pathways in Protein aggregation

The formation of intermediates in the protein folding pathway has high relevance in a protein forming its native structure. The normal folding pathway is known as the non-cooperative or

multistate pathway. In some instances, when the intermediates lie away from the normal pathway, they form “off-pathway intermediates”. Irrespective of these intermediates’ space-time conundrum, they interact, shifting the equilibrium toward forming protein aggregates and acting as a driving source. The intermolecular interactions are much more significant than intramolecular interactions at these stages. The aggregates thus formed are classified as amorphous or ordered. The “amyloid fibril” is the most commonly occurring ordered state and the leading cause of AD (Calamai *et al.*, 2005, Kundu *et al.*, 2020) (Figure 1.3-1.4). Goldberg and his colleagues first developed initial ideas on protein aggregation through a classical experiment using the tryptophanase (tetramer) protein (London *et al.* 1974). They experimented with the unfolding of tryptophanase in 8M urea, followed by diluting the buffer; both refolded and inactive aggregates were formed.

Further, more active enzyme portions formed inactive aggregates when the protein concentration was increased. The group also showed various ways a protein could get trapped between forming a stable native structure and unfolded aggregate state, depending mainly on the denaturant concentration and renaturation step. They found that the entire enzyme was in an aggregated state at the specific concentration of 3M urea, suggesting that it is prone to form some intermediate aggregated state at this particular denaturant concentration. When the group denatured BSA and some crude bacterial extract in the same tube in the presence of tryptophanase, the enzyme had no issues with refolding. This observation made Goldberg and his colleagues believe that the entire self-association process at any particular denaturant concentration is specific. They further suggested that the aggregation pathway has similar stereospecific interactions required by the normal folding pathway. The difference is that when a protein folds into its native structure, many intramolecular interactions get substituted by intermolecular interactions between two polypeptide chains.

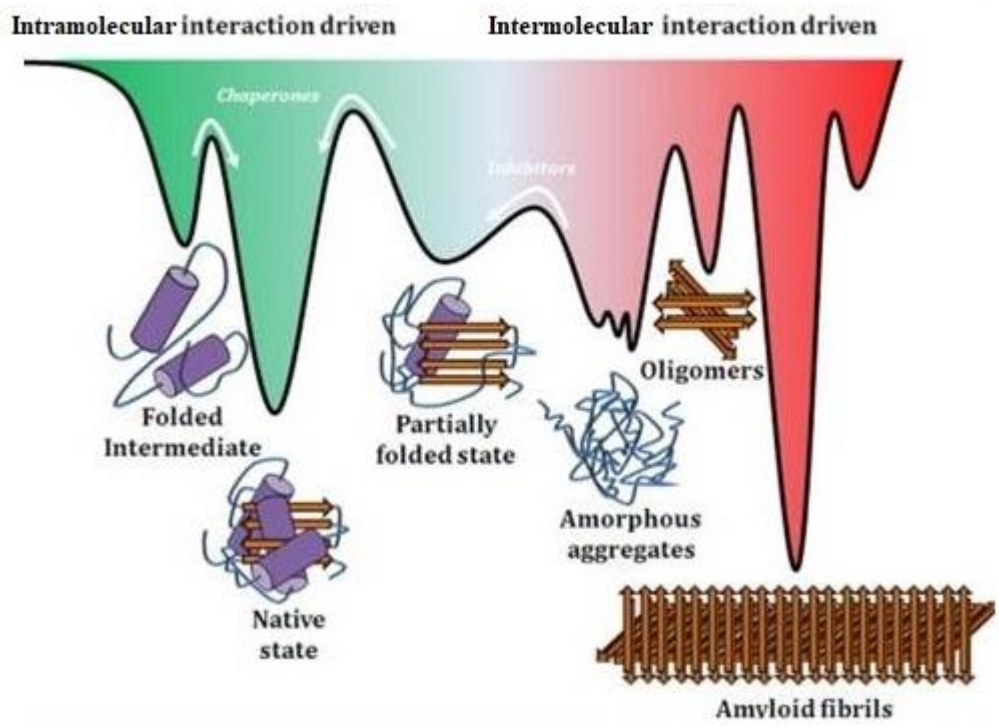

Fig 1.3 The various stages in the protein folding pathway: The formations of protein native states are dominated by intramolecular interactions, whereas the partially folded states followed by the formation of self-assembled aggregates states and oligomers are majorly formed by intermolecular interactions. The amyloid fibrils are energetically more stable structures in the pathway than the native state or the intermediate aggregates (*adapted with permission from*: Denny *et al.*, (2013). Recent developments in targeting protein misfolding disease)

Further, protein conformations with enhanced flexibility and access to solvent are essential for starting the fibrillation process, implying that a partial loss in the folding and the compactness of the structure is a prerequisite for the formation of amyloids (Pedersen *et al.* 2004). Several studies involving many proteins have shown how environmental conditions are essential in forming aggregates and intermediate stages in the protein folding pathway. β_2 microglobulin fibrillogenesis *in vitro* initiates in low pH conditions and high ionic concentrations. There is a distinct presence of unfolded intermediates with significant secondary and non-native tertiary structures (MacParland *et al.*, 2000). In the case of Hen egg-white lysozyme, low pH and high temperature play the role of catalysts in enhancing the formation of aggregates and intermediate stages (Wang *et al.* 2006). Studies have also revealed that these intermediate states illustrate a molten globule-like state which has high surface hydrophobicity, stable secondary structure and non-native conformations, with the hydrophobic interactions playing a significant role in the process of amyloidogenesis (Booth *et al.* 1997, Gerber *et al.* 2007 and Bolognesi *et al.* 2010).

1.4- Cellular mechanisms against protein misfolding

Once synthesized by the ribosome, any protein undergoes a protein folding process in the cytoplasm. The protein folding process begins after the protein is released from the ribosome, which interacts with chaperones and heat shock proteins (Kundu *et al.,* 2020). Efficient translocation of the protein to its functional area is also significant. Within the cell, protein folding could occur after the protein is formed entirely, or sometimes the folding process might also begin simultaneously before the completion of translation when the nascent protein is still attached to the ribosome. This is called co-translational protein folding. Studies in protein folding have shown that the heat shock proteins and molecular chaperones have multifold functions regarding protein folding. They predominantly interact with partially folded proteins and have exposed hydrophobic patches. They help nascent proteins fold correctly, save slightly misfolded proteins, and rescue them, giving them a second chance to undergo the folding process to attain a correctly folded structure. The protein folding process is highly energy-dependent, and it is now understood why ATP is vital for this process. There are also high ATP requirements in localized cellular stress areas (Dobson *et al.,* 2003, Agbas *et al.,* 2018, Kundu *et al.,* 2020). Various heat shock proteins like HSP 40, 70, and 90 assist in the proper protein folding but do not increase the process rate. Thus, various enzymes like protein disulphide isomerase and prolyl peptidyl isomerase have vital functions in protein folding *in vivo* (Sarkar and Dubey, 2010). Prolyl peptidyl isomerase enzyme increases the *cis-trans* isomerization involving proline residues, whereas the Protein disulphide isomerase enzyme helps in the disulphide organization linkages in folding- an important contact in secondary structure formation. The presence of active quality control machinery inside the cell is of absolute importance for the overall governance of the protein folding process. The quality control mechanism which functions properly can easily recognize misfolded proteins and drive them towards extinction, degrading them to prevent further aggregation-related complications. They can also prevent proteins from taking an off-pathway during the folding process and prevent any instance of protein misfolding. In eukaryotes, the quality check of the protein also includes critical processes like glycosylation and deglycosylation. These processes help the cell distinguish between

correctly and incorrectly folded protein before being exported out of the ER through the Golgi complex (Wilson *et al*., 2008).

On the contrary, an ill-functioned quality control mechanism can often lead to protein aggregation and deposition of the aggregates in various organs or tissues (Sarkar and Dubey, 2010, Kundu *et al.,* 2020). Among cell organelles, the endoplasmic reticulum holds the highest importance. The ER is home to various heat shock proteins and protein folding catalysts. This makes the ER the go-to cell component regarding protein folding, and any problems associated with folding protein stem from the endoplasmic reticulum (Kaufman *et al.* 2002, Kundu *et al.* 2020). Protein aggregation also affects the ubiquitin-proteasome system, enhancing the protein misfolding process. The protein misfolding process develops many intermediates, which are directly responsible for the formation of aggregates. With extensive research, these well-established intermediates are now characterized as amyloid fibrils (Dobson, 1999; Chiti *et al.,* 1999).

1.5- Amyloids and associated diseased conditions

Protein misfolding diseases are a class of diseases resulting from any protein failing to fold appropriately and gain its native structure. As a result, these proteins might accumulate in organs and tissues, causing plaques. Accumulation of such plaques in critical organs often causes various pathological conditions. Improper protein folding reduces its biological role and affects its natural production levels (Kundu *et al.,* 2020). Sometimes, the reduced amount of protein formed is also because of the cell's quality control system. Misfolded proteins are subjected to degradation within the cell involving the Golgi-reticulum system. Many diseases occur when a small part of the protein or the parent protein itself converts from soluble structures to highly ordered fibrillar structures, as described in the literature (Chiti and Dobson, 2006). These structures are called amyloid fibrils, and they start accumulating extracellularly in organs or tissues. These ordered fibrillar aggregates can also form intracellular inclusions.

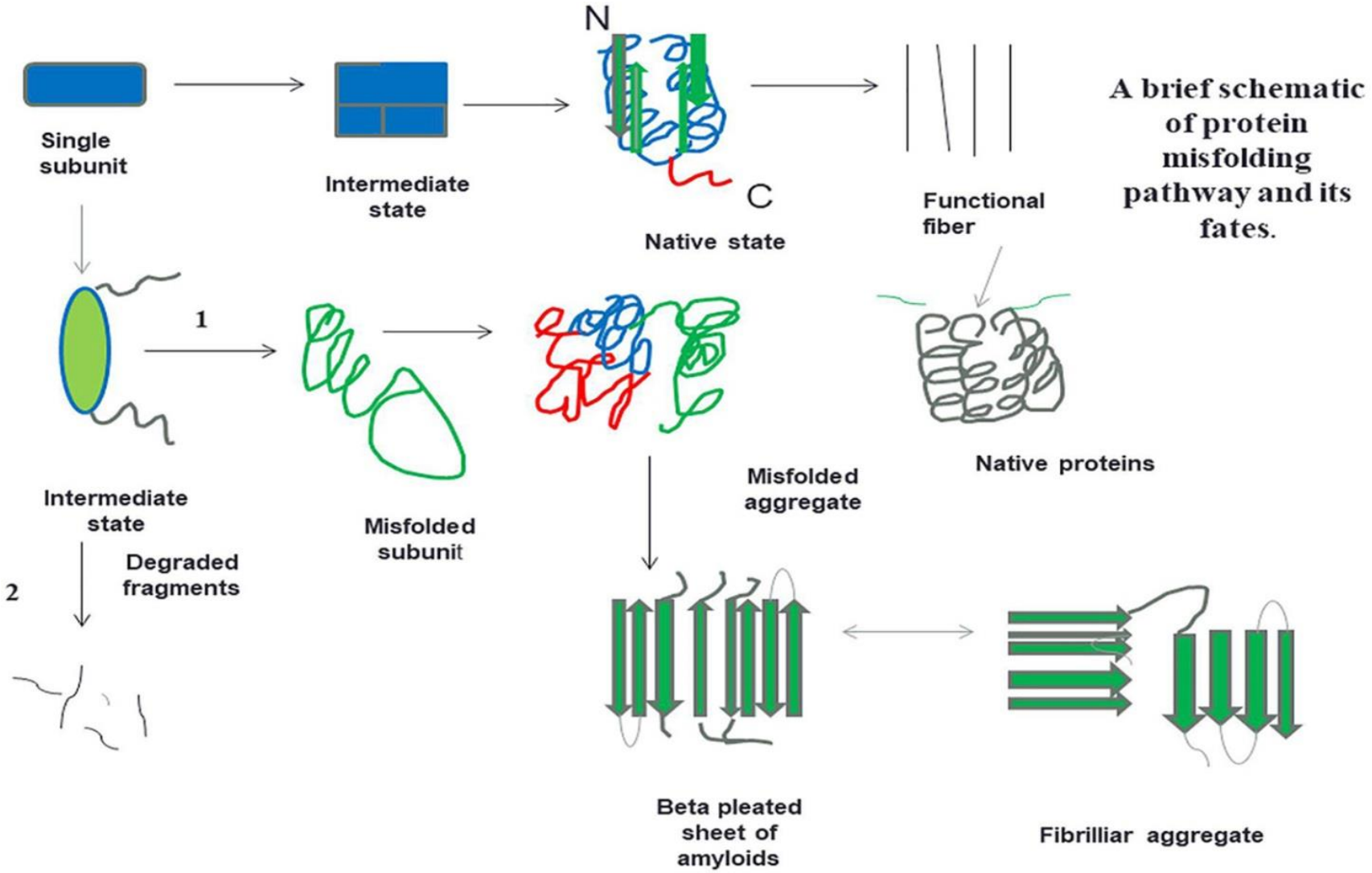

Fig 1.4 Diagrammatic representation of the protein folding process: The polypeptide's various intermediate states can form and take up during the folding process. The illustration also represents the various pathways that could be adopted in protein misfolding and how it could lead to the formation of misfolded subunits, beta-sheet rich regions and finally form, aggregates (Adopted with permission from Kundu *et al*. 2020 Advances in protein misfolding, amyloidosis and its correlation with human diseases)

The difference in the terminologies is based on specific morphological differences between the extracellular and intracellular deposits (Westermark *et al.,* 2005, Ashraf *et al.*, 2014, Kundu *et al.,* 2020). Some protein misfolding diseases could also be due to mutations that shift the amino acid sequences (Ashraf *et al.*, 2014). One of the exciting facts about amyloid proteins is that they are formed from a short stretch of amino acids. There are thirty-seven amyloids associated with pathological conditions that are well-reported and well-established. Most pathological conditions are formed from stretches of 100 amino acids; only four such diseases are known, with more than 400 amino acids and the highest being 700 amino acid residues (Chiti and Dobson, 2017).

Further, most of these proteins fall under both secretory and cytosolic proteins. The secretory proteins usually form extracellular deposits, whereas the cytosolic proteins are more known to form intracellular deposits. Amyloid deposition might occur in the central nervous system, where they can form deposits in specific brain areas, causing neurodegenerative diseases like Alzheimer's Disease (AD) or Parkinson's Disease (PD). The deposition of amyloids in other organs and tissues causes non-neuropathic conditions. Amyloid-related disorders are majorly sporadic and occur in old age. The sporadic nature of amyloid-related diseases mainly suggests that these are majorly caused due to progressive loss of the cell's quality control mechanism due to ageing. A tiny fraction of associated amyloid diseases are hereditary (Dobson, 2003; Powers *et al.,* 2009; Chiti and Dobson, 2017; Kundu *et al.,* 2020). Some amyloid-associated diseases could also arise due to particular medical treatments like amyloidosis associated with dialysis, hemodialysis, and amyloids associated with localized injection sites as a treatment for Type 1 diabetes mellitus and iatrogenic Crutzfelt-Jacob syndrome, which occurs due to organ transplantation. Some cases can also occur from other biological sources, which might be contaminated with prion proteins (Chiti and Dobson, 2017, Kundu *et al.,* 2020). Neurodegenerative conditions like AD and PD are primarily sporadic, although some hereditary form is also reported. The tables below (Tables 1.1 and 1.2) contain the neurodegenerative and systematic amyloidosis lists.

Table 1.1- Amyloid diseases affecting the nervous system of the patients and causing neurodegenerative conditions

Protein / peptide	Associated disease	Native structure
Amyloid β peptide	Alzheimer's disease	Unfolded native structure
Prion protein	Spongiform encephalopathy	Unfolded native structure (1-120) and α helical (121-230)
α Synuclein	Parkinson's and Dementia with Lewy bodies	Unfolded native structure
Tau	Frontotemporal Dementia with Parkinson's	Unfolded native structure
Superoxide dismutase	Amyotrophic Lateral Sclerosis	All β and Immunoglobulin-like fold
Atrophin 1 is associated with Poly Q expansion	Hereditary dentatorubralpallidoluysian atrophy	Not known
Ataxin and Poly Q expansion	Spinocerebellar ataxias	All β-, AXH domain, TBP-like residues
Huntingtin-associated Poly Q expansion	Huntington's disease	Large native unfolded structure
TATA-box- binding protein and poly Q expansion	Spinocerebellar ataxia	Both α +β structure
Androgen receptor with Poly Q expansion	Spinal and bulbar muscular atrophy	α dominant structures, nuclear receptor ligand-binding domain
ABri peptide	Familial British Dementia	Unfolded Native structure
ADan peptide	Familial Danish Dementia	Unfolded native structure

Table 1.2- Non-neuropathic systematic amyloidosis conditions occurring in humans, their associated protein structural details

Protein or peptide	Disease	Structure
light chain immunoglobulin fragments	AL Amyloidosis	All β and immunoglobulin-like
Serum amyloid A protein	AA amyloidosis	All α helical with unrecognized fold
Serum amyloid A protein fragments	Familial Mediterranean fever	All α helical with unrecognized fold
Transthyretin – wild type	Senile systematic amyloidosis	All β and prealbumin-like folds
Transthyretin- mutant type	Familial amyloidotic polyneuropathy	All β and prealbumin-like folds
B2- microglobulin	Haemodialysis related amyloidosis	All β and immunoglobulin-like
N terminal fragments of Apo AI, AII and AIV	ApoAI, ApoAII and ApoAIV amyloidosis	Unknown
Gelsolin mutant fragments	Finnish hereditary amyloidosis	Natively unfolded structure
Lysozyme mutants	Lysozyme amyloidosis	α + β folds, lysozyme fold
Fibrinogen variants	Fibrinogen amyloidosis	Unknown
Cystatin C	Icelandic hereditary cerebral amyloid angiopathy	α + β folds, cystatin fold
N terminal fragments of Apo IV	Apo IV amyloidosis	Unknown

1.6 Amyloid formation mechanism

(A) From Amyloid Precursor Protein to Amyloid β- sheet formation*:*

Amyloid precursor protein (APP) is a type I integral membrane protein of approximately 770 amino acids. The Aβ region, which forms the plaques, is located towards the N terminal end of the APP. APP can be subjected to sequential cleavages in a non-amyloidogenic pathway by α-secretase (ADAM 10) in the ectodomain, forming larger soluble fragments instead of insoluble Aβ fragments. The action of this enzyme also releases the soluble APPα portion from the N-terminal of the protein. A small CTF83 fragment is next subjected to cleavage by γ secretase. The action of γ secretase produces small fragments of p3 (Aβ 40/42), which have low cellular toxicity. The intracellular domain of the APP protein (AICD) is released inside the cell. In the amyloidogenic pathway, the APP protein is subjected to cleavage by β-secretase (BACE 1 or β-site APP Cleaving Enzyme). This enzyme also produces a soluble APPβ (sAPPβ) and CTFβ-99. γ secretase acts upon CTFβ-99, further producing toxic forms of Aβ40/42. These forms are then transported to the cell surface and secreted by recycling vesicles (Figure 1.6). Extensive research has further established that the long form of Aβ, i.e. Aβ42, is more toxic than Aβ40 (Carillo-Mora *et al.*, 2014, Ashraf *et al.*, 2014).

It is also essential to understand that inherently APP or Amyloid β are not harmful in their native state in the human body. The fundamental physiological roles of amyloid β peptides are yet to be deciphered fully. Some areas anticipated to have some roles include kinase activation, protection from oxidation of metals, cholesterol transport and ion channel damages (Maloney and Lahiri, 2011; Yang *et al.*, 2014; Kundu *et al.*, 2020). Also, the amyloid precursor protein (APP) is expressed in all mammalian tissues. Specifically, the amyloid precursor protein has high expression in kidneys and brains. Lower organisms like *Drosophila melanogaster* and *Caenorhabditis elegans* also express similar proteins suggesting that this protein is also functional during these organism's developmental stages. *In vitro* studies have further confirmed that APP cleavage products also have a role in cell adhesion and the growth of neurons (Yang *et al.*, 2014).

(B) Formation of amyloids *in vitro*:

Under laboratory conditions, scientists must replicate conditions that enhance the formation of oligomers and amyloids from various source proteins. Understanding what physicochemical factors might govern the formation of amyloids is essential. The key in the experimental environment is to identify the key stages that any polypeptide undergoes during the formation of amyloids. Biophysical experiments have shown that amyloid fibril formation is preceded by forming different monomeric and oligomeric conformational states. The role of the protein's primary structure in such cases is also essential. There is increasing research in understanding the roles of these oligomeric species that precede the formation of amyloid fibrils which are now understood to have a significant role in amyloid pathology and toxicity. The most widely accepted hypothesis for the formation of amyloids in experimental conditions is a *nucleated growth mechanism.*

The growth curve of amyloid fibrils or oligomer formation is sigmoidal. A lag phase represents the time required to form the nucleus, which comprises a threshold number of molecules. Once this nucleus is formed, the lag phase accelerates towards an exponential phase. The exponential phase is characterized by adding monomeric or oligomeric species to this initial nucleus formed (Chiti and Dobson, 2017). After the exponential phase, the amyloid fibril formation's final stage is the stationary phase, corresponding to an entire fibril formation where no further elongation occurs. The lag phase is a variable component in the formation of amyloid fibrils. It depends on the nature of the model protein and the specific aggregation employed. Reports suggest that certain mutations reduce the process of nucleation; hence lag phase appears to be lacking before the elongation phase (Uversky *et al.* 2002; Pedersen *et al.* 2004). The absence of a lag phase does not mean that the nucleation step is absent. It only implies that the rate of formation of fibrils supersedes the rate of formation of the nucleus. The nucleation rate does not act as the limiting step in the formation of fibrils. Further, in some cases, the nucleus forming phase is replaced or shortened by adding some preformed fibrils (oligomers/monomers) under the experimental conditions. This process is called *seeding.* Although, the ability of any preformed fibrils to

accelerate the formation rate of additional fibrils depends on the sequence and is reported to lower the rate of formation of fibrils dramatically because of divergence in sequence (Wright *et al.* 2005) (Figure 1.5).

Before forming matured fibrils, a more presumed toxic, spherical oligomeric species are usually formed. These toxic and spherical structures are called protofibrils and have a substantial role in amyloid-associated pathogenicity. They can exist as independent spherical-shaped structures of 2-5 nm diameter or stay associated with forming a linear chain-like structure (Chiti and Dobson, 2006). These protofibrils have a characteristic β sheet structure and other structural features that bind to classical dyes like ThT and Congo red (CR).

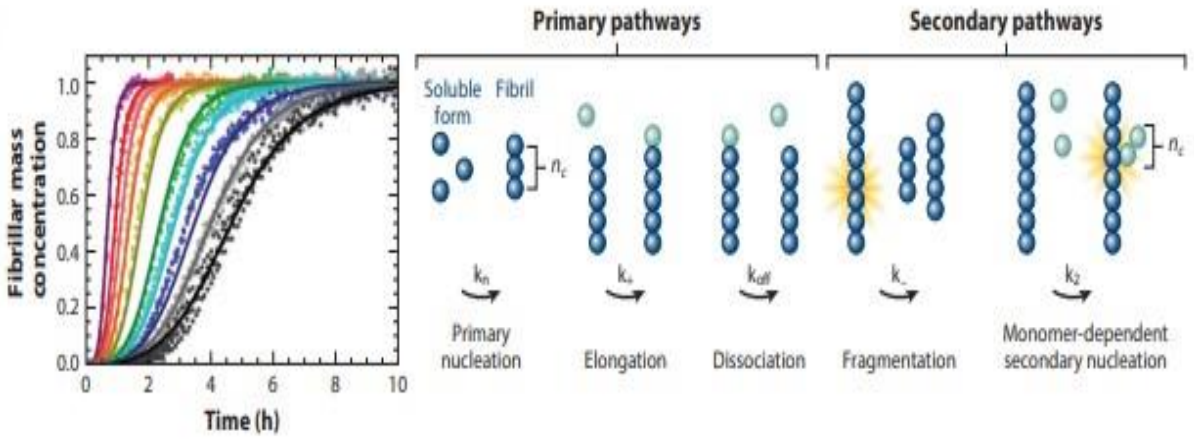

Fig 1.5 Representation of amyloid formation kinetics: Classical sigmoidal kinetics, which is typical of the protein aggregation process. Further, the primary and secondary molecular pathways dominate protein aggregation. Primary nucleation followed by elongation and dissociation (fragmentation) followed by further elongation depending on secondary nucleation (*adopted with permission from:* Chiti, F and Dobson, M.C. (2016) Protein misfolding, amyloid formation and human disease: A summary of progress over the last decade).

Literature also suggests that these protofibrils have more ThT or CR binding activities than matured fibrils and more exposed hydrophobic patches (Bolognesi *et al.* 2010). Some specific antibodies bind exclusively to these protofibrillar species but not to the standard linear protein or the matured fibrils (Kayed *et al.*, 2003). This observation is important because this specific state of amyloid formation should be targeted to develop drugs against neurodegenerative diseases.

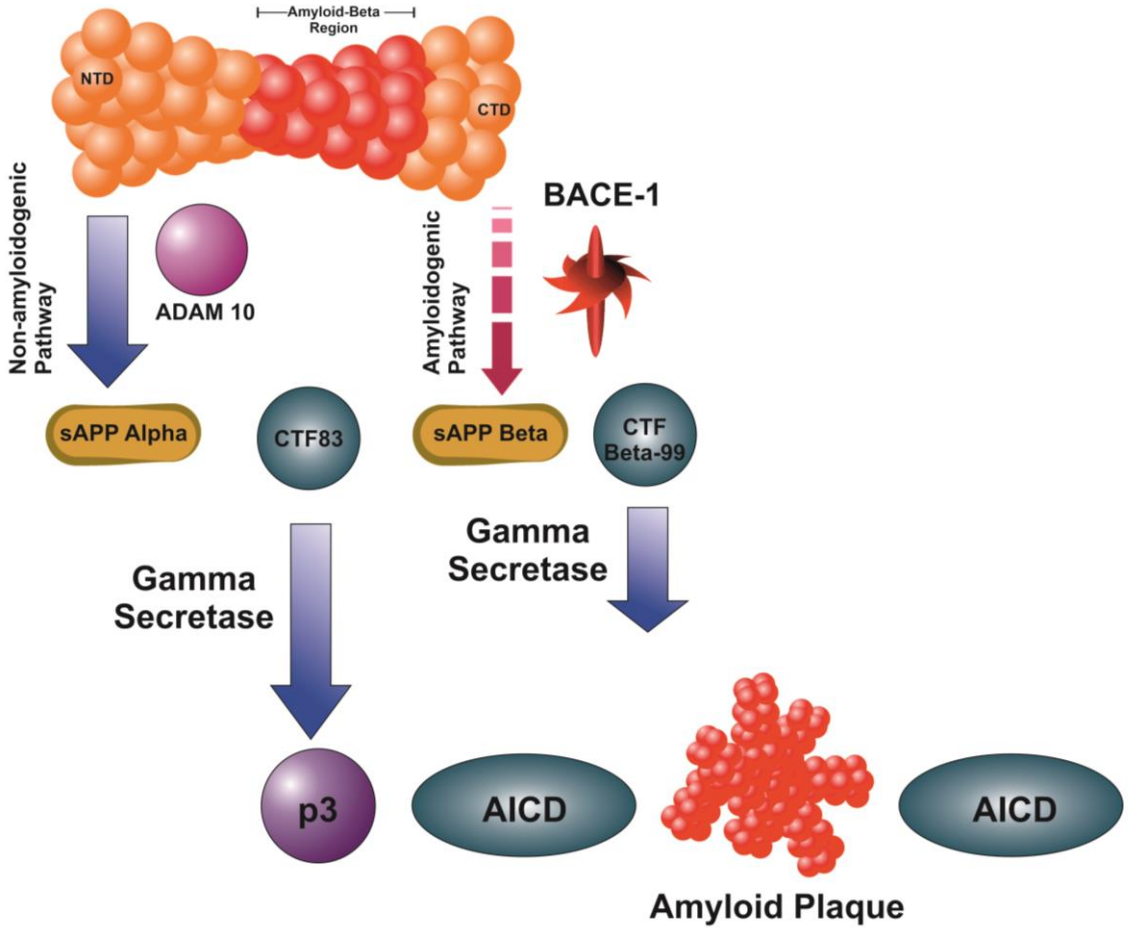

Fig 1.6 Diagram showing two proteolytic pathways of the amyloidogenic precursor protein. ED and ID are extracellular and intracellular domains, respectively. APP is a membrane-bound type-1 protein processed by two pathways, the amyloidogenic and non-amyloidogenic pathways. In the non-amyloidogenic pathway, the APP's proteolysis in the ectodomain of the protein produces more extensive soluble Aβ fragments. This step is catalysed by the α Secretase. In addition, the protein's C-terminal domain is attached to the membrane and processed by γ Secretase, which produces smaller fragments with much less cytotoxicity called p3 and the intracellular domain of APP, which is known to show specific neuroprotective activity. In the amyloidogenic pathway, the enzyme that causes proteolysis of the APP in the N terminal, APPβ and CTF β. CTFβ is further processed by γ secretase, which produces Aβ fragments insoluble. Gamma secretase plays a critical role in the pathway.

Certain other experimental conditions also help us enhance the probability of the unfolding of the protein, enhancing the aggregation rate. Before it goes into the state of aggregation, every protein first attains a partially unfolded state with extreme experimental conditions like high temperature, extreme pH, high concentration of the protein, or presence of certain organic solvents that help the native protein get to that state. One key factor in forming amyloids under *in vitro* conditions is destabilizing the protein's native structure. The destabilization of the native structure increases the amyloid-forming propensity (Chiti *et al.,* 2000). The partially unfolded state of the protein has structural characteristics where the protein is void of its tertiary structure characteristics but has the secondary structure almost in its original form. The secondary structure also has high surface hydrophobicity. The aggregation process starts from the exposed β domain in the monomeric intermediates (Selkoe, 2003). The monomeric states of the proteins undergo a structural transition where β-sheet structures overtake the available ordered secondary structure. It is now well established and understood that fibrillation only proceeds when this prefibrillar state is formed from partially unfolded structures with higher amyloid dyes binding capacity and exposed hydrophobic surfaces.

1.7- Structural characteristics of amyloids

The molecular structure of amyloids has now helped us understand how monomeric proteins interact with each other, which is an essential step in forming amyloids. The importance of amyloid structural studies is because understanding the step-by-step processes and the exact structural transitions will help scientists develop drugs more effectively, targeting critical stages in the formation of amyloids. In an attempt to unravel the amyloid's core characteristic feature, the first X-ray diffraction studies were made in 1997. It revealed a helical array of β sheets and long strands running parallel to sheets perpendicular to the axis (Blake, 1997). Other methods of determining the presence of amyloids like dyes and biophysical methods like Atomic Force Microscopy (AFM), Scanning Electron Microscope (SEM), and Transmission Electron Microscope (TEM) were on the rise (Khan *et al.* 2012; Kundu *et al.* 2020). Techniques like fluorescence spectroscopy, magnetic resonance spectroscopy, single-photon emission computed tomography and multiphoton

microscopy are now employed (Chiti and Dobson, 2017). Thioflavin T is also a very sensitive assay for detecting amyloid *in-vitro*. ThT can bind more aggressively with fully matured fibrils, but in the case of amorphous aggregates, it is highly dependent on pH, and slight changes in pH can induce micelles formation by ThT, decreasing the ThT intensity multifold and giving erroneous readings (Khurrana *et al.* 2005; Khan *et al.* 2017, Kundu *et al.* 2020). The advent of solid-state NMR (SS-NMR) helped better characterize the structural features. Tycko and colleagues deciphered the amyloid β peptide's core structure, the use of ss-NMR, and the computational energy minimization process. The experiment was done at room temperature and physiological pH of 7.4. The experiment revealed the core structure of Aβ 1-40, which showed a parallel arrangement of β strands involving the region between the 12^{th}-24^{th} residue and the 30^{th}-40^{th} residue forming the fibrils. Another essential feature of this arrangement is the loop between the 25th-29th residue connecting the two β-strands that do not belong to the same β sheet but are two distinct β sheets of the same protofilaments.

Further research showed that different Aβ molecules form a stack in a parallel arrangement. Each protofilament comprises four β sheets separated by approximately 10Å (Petkova *et al.,* 2002; Chiti and Dobson, 2006). Other than amyloid β peptide, essential studies were carried out in other proteins that form amyloids, like transthyretin and a small hexapeptide (GNNQQNY) in Sup35 *et cetera.* Detailed studies using ss-NMR, X-ray crystallography and cryo-EM have given us more information on the Aβ protofilaments. Various structural models established that the protofilaments are 6 nm wide, having lateral associations forming a striated ribbon-like structure. There have been contradictory reports on this structure where Paravastu *et al.* reported that the protofilaments were 7 nm wide and lacked striated ribbon-like structures (Paravastu *et al.* 2008). Besides the standard features reported, there is a dichotomy with protofilaments' striations. The new study also revealed the second polymorph in the arrangement of the β sheets. The second polymorph revelated that instead of four β sheets, protofilaments could also consist of six β sheets in a threefold symmetry. Critical side-chain interactions are also different in different protofilaments. Later studies using

similar techniques have revealed that more polymorphs could exist within the same polypeptide (Colvin *et al*., 2016; Chiti and Dobson, 2017). Similar techniques for other proteins include human IAPP, human PrP, human insulin, and β microglobulin, although the protofilaments' exact structure is yet to be worked out fully (Helmus *et al*., 2011, Yang *et al.,* 2010, Bateman *et al.,* 2010, Chiti and Dobson, 2017).

X-ray diffraction studies have been critical in revealing various aspects of amyloids' structure and their various polymorphisms. The studies have usually been carried out in small peptide segments of no more than ten peptides, and this model has been used to reveal the various standard features of any amyloid system. Each peptide molecule within the system forms β-strands, and multiple β strands form in-register alignment, further forming β-sheet. The alignment of the β strands supports favourable interactions, which drive the formation of parallel β-sheet arrangements. Optimal side-chain interactions and parallel β sheet arrangements are vital factors that drive the aggregation process. The polymorphism of the amyloid structures is classified based on various factors:

1) the arrangement of the β strands in the sheet (parallel or anti-parallel)

2) interaction pattern of adjacent sheets being either in face-to-face orientation or face-to-back orientation

3) the direction at which the strand termini of different sheets points either in the same direction or opposite.

The various polymorphisms of the amyloid structures can be well studied using an electron microscope. Various structural features of amyloids include the degree of the twist, the number of filaments in each fibril, the diameter, and the mass per unit length (Eisenberg and Jucker, 2012; Chiti and Dobson, 2017; Riek *et al.,* 2017). The different conformational states of the amyloids and their interconversions are highly dependent on the local cellular environment. The various intermediate states that a protein shows in protein folding and early stages of amyloid formation are

essential and utilized by the cell in various stages of any biological process. Failure of the cellular systems to keep these processes in check often leads to metabolic disorders.

1.8- Factors modulating the amyloid formation

All proteins have an inherent property of behaving like amyloid and forming aggregates, although it depends much on the local microenvironment. Classical biochemistry studies also suggest that the sequence of the native also plays an essential role. The protein sequence determines which amino acids are more accessible to the solvent and which amino acids stay in the core during protein folding. This process somehow controls the protein-forming amyloids. Different proteins with linear structures might form amyloids at different rates, mainly dependent on the primary sequence. The proteins will have a common starting point: a partially folded structure and an improperly folded native state.

(A) Hydrophobicity and secondary structure: The number of hydrophobic amino acid residues in a protein is an essential factor in forming amyloids by any protein. Substitution of amino acids, which are critical for the proper folding process of that protein, or in the area of the protein that has a vital role in folding, substitution by hydrophobic amino factors increases the formation of amyloid-forming propensity. Similarly, replacing hydrophobic residues with hydrophilic residues decreases the propensity for the same (Chiti and Dobson, 2006). Studies have further shown that the natural evolution of proteins has led them to avoid having consecutive groups of hydrophobic amino acid clusters, increasing amyloid-forming propensity. Instead, evolution has favoured proteins that have successfully avoided such circumstances (Chiti and Dobson, 2006).

The net charge on a protein also has a role in increasing the propensity for amyloid formation. Any protein or peptide sequence having a high positive net charge has less propensity for self-association. Studies on the protein AcP have shown that mutations that decreased the overall charge on the protein accelerated the formation of β sheets, whereas the mutations which increased the overall net charge on the protein decreased the formation of β sheets. The increase in

β sheet content was validated by binding with amyloid binding dyes like ThT and Congo Red (Konno, 2001). Computational investigations also suggested that proteins that form partially unfolded intermediate states have lower hydrophobicity and higher overall net charge than those that do not form intermediate states (Chiti and Dobson, 2006). This combination of hydrophobicity and net charge allows the intermediates states to show lower or avoid the formation of aggregates in normal physiological conditions despite most of the side chains of various amino acids available for intermolecular interactions. Evolution has also prevented proteins from forming sequences that do not have alternating hydrophobic or hydrophilic residues. This sequence stabilizes the formation and interactions of β-strands, increasing the probability of forming β-sheets (Broome and Hecht, 2000).

(B) Sequence of Amino Acids: Experiments in the AcP model protein where various amino acid substitutions were carried out showed that mutations impact the protein's aggregation forming capacity. The experiments revealed certain factors that led scientists to develop a model equation to calculate the aggregation rate, taking various extrinsic and intrinsic factors into account. The equation was developed based on rigorous experimental results. This model also helped us differentiate between mechanisms and similarities that govern amyloid formation in different proteins. Other factors include the number of aromatic side chains, net dipole moment, and exposed surface area of amino acids. These factors are now also included in estimating amyloid-forming propensity. These factors help us understand when a protein can behave as a polymer and behave in a pro-aggregator way compared to the standard folding process of globular proteins that are majorly influenced by the structures present in the native state of the protein. The simple understanding of the various factors that influence the formation of β sheets has led bioinformaticians to develop softwares that could predict the amyloid-prone regions in any peptide or protein sequence. The software developed has been rigorously used to predict the forming aggregation zones in standard protein and validated by experimental data. The protein used for validation included amyloid β peptide and α- synuclein protein (Pawar *et al.,* 2005; Chiti and Dobson, 2006).

Disulphide bond formation is also a relevant factor that influences amyloid formation. Literature reports various disulphide bonded globular proteins which form amyloids under specific conditions. Studies using reducing agents such as DTT and cysteine inhibited the formation of amyloids in β2 microglobulin. β2 microglobulin causes dialysis-associated amyloidosis in humans (Yamamoto *et al.* 2008). Other studies which support the hypothesis of inhibitory effects of reducing agents towards amyloidogenesis include reports using Tris (2-carboxyethyl) phosphine in reducing amyloidogenesis in hen egg-white lysozyme by breaking of disulphide bonds and initiating protein unfolding (Wang *et al*. 2009) and the role of DTT in reducing amyloidosis in HEWL at alkaline pH (12.2) (Kumar *et al.* 2008). Such reports enhance the possibility of designing drugs that affect the disulphide bonds as anti-amyloidogenic agents.

(C) Unfolded areas in the protein: Other than the key factors discussed above that are established as factors running the possibility of forming amyloids in any protein, the actual aggregation rate is also highly governed by the wide range of secondary structure interactions. Studies on model AcP with limited proteolysis in the presence of trifluoroethanol in moderate amounts showed that initial aggregation-prone regions were further solvent-exposed, further enhancing the amyloid-forming capacity. Other regions with a higher amyloid-forming propensity were partially buried inside the core, reducing their chances for amyloid formation (Monti *et al.* 2004).

1.9- Nucleotide imbalance and neurodegeneration

Purines are double-ring organic molecules that are heterocyclic and aromatic. The primary purines are adenine and guanine, whereas pyrimidines, on the other hand, are singled ring molecules. They are cytosine, thymine and uracil. The purines and pyrimidines are the most critical components of RNA and DNA. Apart from the nucleic acids, they also form nucleosides and nucleotides. The nucleotides are formed by adding one, two or three phosphate groups associated with the purinergic nucleosides- adenosine, guanosine, and pyrimidinergic nucleosides- cytidine, uridine and thymidine. Other than these conventional molecules, altered chemical modifications

form altered metabolites like xanthine and hypoxanthine (Ansolega *et al.,* 2014). Adenine is an integral part of co-enzymes and cofactors and an essential enzymatic component. Specific adenine-associated cofactors include nicotinamide adenine dinucleotide, flavin adenine dinucleotide, and nicotinamide adenine dinucleotide phosphate in their reduced forms. Purines and pyrimidines are essential components of nucleic acids and their synthesis, but they are also crucial as they provide metabolic signals, energy regulating cell growth, and are part of co-enzymes. They further modulate sugar transport. Pyrimidines are also important as they regulate the production of phospholipids and polysaccharides, glycosylation of protein and lipid and other detoxification processes (Fumagalli *et al.,* 2017). Nucleotides are formed from purinergic and pyrimidinergic nucleosides after adding phosphate moieties. The liver usually synthesizes nucleosides *de novo* (Figure 1.7).

They are also well obtained through various food sources in our diet. After the nucleosides are formed, they are transported to the brain by nucleoside transporters, where they are converted to their final form, their respective nucleotides intracellularly. There is usually significantly less synthesis of nucleosides in the brain. These nucleosides' balance and successive generation are usually maintained by de novo synthesis and breakdown from other metabolites by critical biochemical pathways. The pathway is called the Salvage pathway. Intracellular degeneration of nucleotide triphosphates like ATP and nucleotide diphosphates (ADP) are broken down to their respective monophosphate form (AMP) by triphosphate phosphatases and diphosphate phosphatases. AMP is converted to IMP (Inosine monophosphate) through the IMP pathway and adenosine by the AMP deaminase (AMPDA). Whenever a cell divides, the resources for stable genome replication must always be fulfilled. ATP is a molecule equally crucial to both dividing and quiescent cells. ATP is also critical for neural tissues produced by continuously functioning mitochondria. Due to mitochondria's continuous activity it is often subject to oxidative damage causing stress to DNA replication (mitochondrial DNA). Oxidative stress can occur due to a loss of balance in the ratio of nucleotides and affects the replication of the mitochondria genome. This kind of inborn error in metabolism can often lead to infantile death, although there are reported

variations in these diseases' actual penetration and clinical manifestations. As a result, maintaining the correct ratio of nucleotides is crucial for maintaining genome replication integrity. This ratio is maintained by these two pathways mentioned above (Fig 1.4) (Micheli *et al.,* 2011; Fasullo and Endres, 2015). Improper replication of DNA results in increased production of dNTPs used to compensate for DNA replication and repair (Chabes *et al.*, 2003, Fasullo and Endres, 2015). Certain single gene defects often affect these critical biochemical pathways, ultimately leading to the untimely development of various pathological conditions affecting the nervous system (Fu *et al.*, 2014; Fasullo and Endres, 2015).

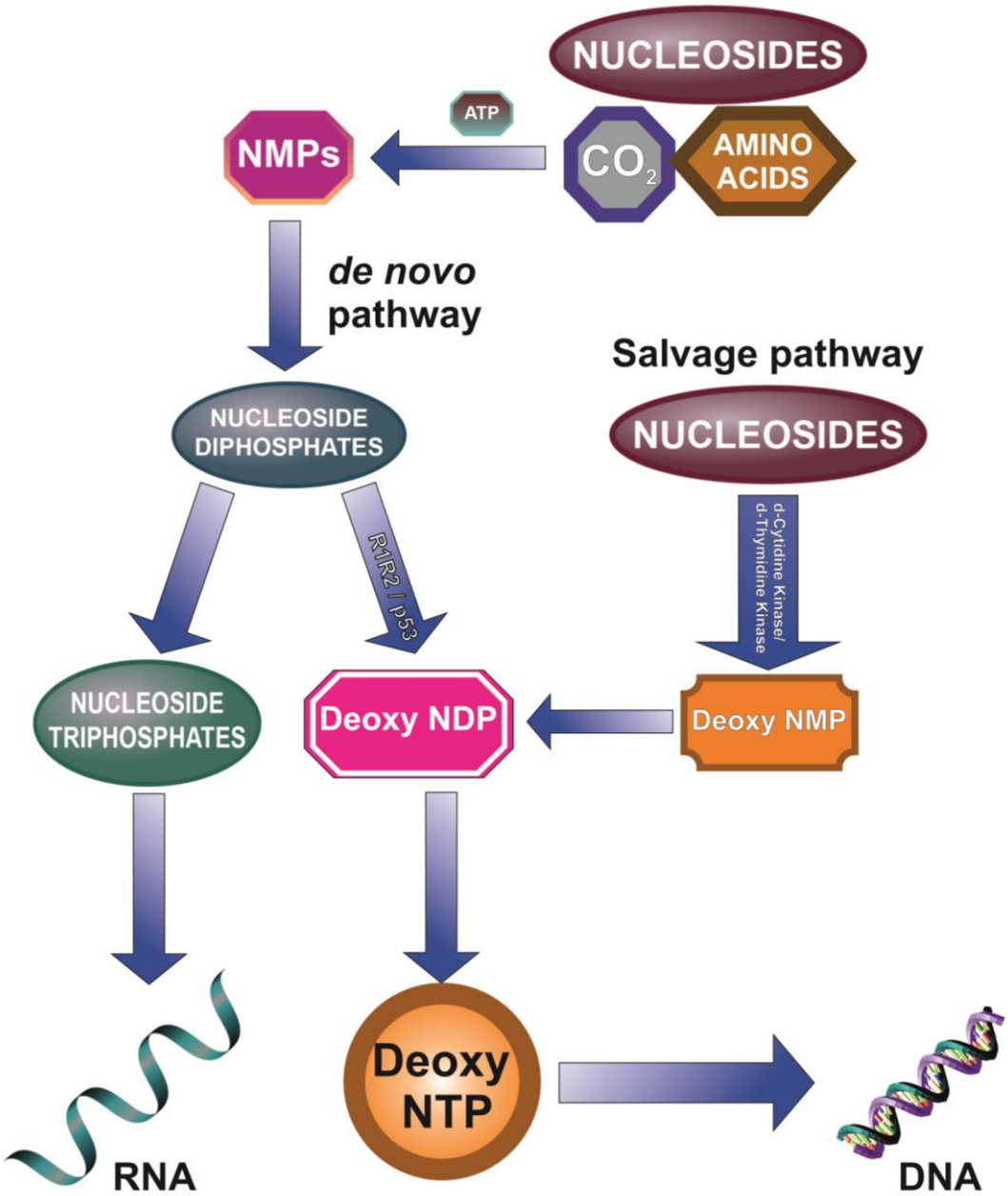

Fig 1.7 Illustration of the *de novo* pathway highlights the incorporation of nucleoside triphosphates (NTPs) and dNTPs in RNA and DNA. Nucleosides, carbon dioxide and amino acids are the raw materials where a phosphate group from the ATP is added, forming the nucleoside monophosphate. Through the *de novo* pathway, nucleoside diphosphate is reduced to deoxy form (d-NDPs) by the action of ribonucleotide reductase large and small subunits (R1R2) and p53 induced ribonucleotide reductase small subunit. d-NDPs are also formed from accumulating nucleosides (recycling) when they are first converted to their monophosphate form by cytosolic enzymes like thymidine kinase-1 and cytidine kinase-1, which are later converted to dNTPs for incorporation into DNA. This recycling pathway is called the salvage pathway. The *de novo* pathway is an energy-dependent process. The R1R2 and p53 inducible R2 activity is highest in the transcriptional and translational stages; the highest activity is reported in the S-phase of the cell cycle in which the genome is replicated to produce sufficient dNTPs.

(A) Purinergic Signaling: After the initial reports in the 1970s that ATP could act as a neurotransmitter in noradrenergic and noncholinergic nerves supplying to the gut, it did not receive much recognition. Later on, extensive studies revealed a unique signalling method called purinergic signalling. Purinergic signalling came first into the foray in 1976 when Burnstock described the class of receptors P1 (Adenosine/ATP) and P2 (ADP). There are currently seven established P2X receptor subtypes and eight P2Y receptor subtypes, including some responsive receptors to purines and pyrimidines. Purinergic signalling has multiple roles. In light of current research, some roles include neurotransmission, neuromodulation, chemoattraction, inflammation, trophic signalling, proliferation, differentiation, motility, development and regeneration, and wound healing (Burnstock, 2018; Kundu and Dubey, 2020). Purinergic signalling is also essential for other functions in the nervous system. The functions chiefly include stem cell differentiation, interactions between glial cells and neurons, diseases of the nervous system, neurovascular and neuroimmune connections, plasticity and synaptic transmission (Newman, 2003; Burnstock, 2003).

(B) Nucleosides, Nucleotides and Neurodegenerative diseases: Among various components in the ecosystem of nucleosides and nitrogenous bases present in the brain, the distribution of nitrogenous bases and other vital metabolites and enzymes is invariable among various parts of the brain. The difference in the distribution percentages hints that the distribution pattern depends on particular brain parts where the function of that metabolite is most relevant. The functions of adenosine and non-adenosine-based nucleosides are now well-established, stating that they have functions associated with ideal physiological conditions and under diseased states (Kovacs *et al.*, 2013). The primary physiological functions that adenosine and non-adenosine-based nucleosides regulate include a regular sleep cycle. Further, they also play a role in memory and pain sensation. They usually have their role in developing depression, schizophrenia, epilepsy, and other neurodegenerative diseases under a diseased state. Under these ill conditions, these critical metabolites have a definite imbalance. This confirms that the metabolites, essential enzymes, receptors and transporters are closely knit in maintaining proper functions. Extensive reviews have

previously mentioned that 5' exo-nucleotidase is an essential enzyme that could be targeted therapeutically. This enzyme catalyzes the breakdown of nucleoside monophosphates into nucleoside and inorganic phosphates. This particular enzymatic action helps regulate and maintain internal nucleoside levels, ensuring their respective transporters and receptors properly carry out their functions. A separate class of nucleotidases exists for those functions exclusively at the synapses. These enzymes terminate the nucleotidases' activities in their respective receptors at the synapses and regulate the critical metabolites (Grazia, 2011; Kovacs *et al.*, 2013).

Adenosine: Extensive studies have now reported that adenosine has potential roles in various neuromodulatory activities, and as a consequence, loss of balance is often observed in diseased states. Adenosine is found endogenously in both neurons and glia. It performs various functions, including neuronal transmission (synaptic transmission) and neuronal excitability within the nervous system. G-protein coupled receptors A1, A2A, A2B, and A3 are the primary receptors for such functions. Adenosine functions to regulate sleep, arousal, cognitive functions, memory, and neuronal sleep properly.

Adenosine Triphosphate (ATP): ATP, the ultimate cellular energy source, is required by active and quiescent cells. A continuous supply of ATP is essential for neuronal cells to function correctly. For the continuous supply of ATP, replication of the mitochondrial genome maintains sound fidelity. Since mitochondria are prone to oxidative stress, nucleotide imbalance could lead to faulty genome replication. Loss of nucleotide imbalance is a concern in causing certain inborn genetic defects among infants, which could be fatal. Such loss of nucleotide imbalance could also cause mitochondrial and chromosomal instability in higher eukaryotic organisms (Ikeda *et al.*, 1993; Fasullo and Endres, 2015).

ATP stimulates astrogliosis, hypertrophic and hyperplastic responses, and various trophic factors. In the early '90s, for drug development in the case of AD and PD, clinical trials had some purine derivatives (propentofylline and AIT-082) that had shown enhanced production of

neurotrophic factors in the brain and spinal cord. AIT-082 also had potential anti-neurodegenerative properties and showed reversing action against age-induced dementia in mice models. On the other hand, ATP is reported to enhance neurotransmitter GABA release from hippocampal cells and simultaneously inhibit the excitatory neurotransmitter Glutamate's release. This action shows the protective nature of ATP (Burnstock, 2006; Grazia, 2011; Kundu *et al.*, 2020).

Guanine-based purines (GBPs): Recent research in guanine-based purines has established vital possible roles of guanine and other GBPs as metabolites and signalling molecules. They are hypothesized as critical extracellular molecules playing an essential role in the central nervous system's proper physiological functions. GBP(s) is reported to play some role in neuroprotection, behavioural responses and neuronal plasticity. However, the exact mechanism that underlies these functions to be carried out correctly is still not understood. Also, GPBs do not have any exclusive receptors of their own. Recent literature also suggests that GBPs work closely with the adenosine receptors. These observations enhanced discussions about how adenine-based and guanine-based purines could work so intricately. Guanine-based purines, including GMP and GTP, strongly correlate with properly functioning G receptors.

Further, adenine-based purines also modulate cell differentiation, neuronal plasticity, and overall nervous system survival. However, the importance of research lies in the fact that GBPs are still at a very nascent stage despite these known facts and are called infant neuromodulators. GTP and ATP are usually stored in storage vesicles and are released by exocytosis when required. They are subjected to the same enzymatic action of the 5' exo-nucleotidases—one of the critical differences between ABPs and GBPs. Adenosine is constitutively released from neurons, glial cells, and astrocytes, whereas Guanosine and inosine are derived from their parent molecules. Research has also led other important facts to come to light. It is now known that both GBPs and ABPs are present simultaneously in the extracellular region and are present in a constant ratio despite their differences in their rate of metabolism, mechanism of action, and affinity for their transporters and receptors in the transmembrane system.

The concentration of GPBs is also enhanced after ischemic injury to the brain than their ABPs counterparts (Giuliani *et al.* 2012; Stentoft *et al.* 2014; Di Liberto *et al.* 2016). Recent studies have also suggested that guanine-based purines are neuroprotective in hypoxia and ischemic conditions. These effects have been majorly studied in various mice models *in vivo*. Guanine-based purines are also protective against Glutamate toxicity. Glutamate toxicity often occurs when glutamate is released from neurons and glial cells. Release of glutamate, depolarization of membrane of the neuron, increases the intracellular calcium concentration, ultimately leading to loss of synaptic functions and cell death. Guanine-based purines also regulate the NMDA-induced neurotoxicity in neocortical and hippocampal cells. They also modulate any alterations due to increased proline-like activity in glutamatergic homeostasis because of reduced glutamate uptake. Further, any alterations in the sodium-potassium ATPase activity are also monitored. Such alterations increase intracellular ATP concentration. Lastly, they also have been shown to reduce the effects of convulsants and other effects of epilepsy in various seizure models (Torres *et al.* 2010; Ferreira *et al.* 2012; Brassai *et al.* 2015; Kovacs *et al.* 2015; Di Liberto *et al.* 2016).

(C) Pyrimidines and Neurodegeneration:

In a similar method to their purine counterparts, pyrimidines and pyrimidine-based nucleotides like Uridine triphosphate (UTP), Cytidine triphosphate (CTP) and Thymidine triphosphate (TTP) are metabolized intracellularly. They are also reduced to their respective nucleosides and nucleotides: Uridine, Cytidine and Thymidine. Dihydrouracil and dihydrothymidine are the end products of pyrimidine nucleotide's normal metabolism (Kovacs *et al.,* 2013, Kundu and Dubey, 2020). Purines and pyrimidines both are required for ideal physiological conditions to function correctly. Pyrimidine nucleosides and nucleotides and their purine counterparts are essential for cellular biochemical reactions as much as they are essential as building blocks of nucleic acids. They provide essential sugar moieties for glycogen, glycoprotein, and phospholipids, which are cell membrane components. These molecules are also important in cell signalling. Cytidine phosphate stimulates and regulates phosphatide formation in neurons of the

brain (Conolly and Duley, 1999). On a similar note, Uridine nucleotides are also crucial for various functions of pyrimidinergic receptors: some of the essential functions they supervise include phagocytosis of microglia, establishing communications between glia and neurons most importantly, playing a role in the release of neurotransmitters from presynaptic vesicles (Figure 1.8). Table 1.3 represents various conditions associated with pyrimidine nucleotides and nucleosides and the impact of their imbalance caused due to any inherent mutation of their gene. Further, Table 4 represents certain neurological conditions in which purines and pyrimidine advance in pharmaceutical and drug development (Kundu *et al.*, 2020).

D) Drug development against amyloid diseases: Current and future strategies

Many approaches are in place to develop drugs against various types of amyloid diseases. The most widely occurring amyloid-based disease is Alzheimer's disease (AD). The underlying mechanism of this disease is still not clearly outlined in the literature, and extensive research is still going on to investigate the true nature of the disease. Extensive research reports generated have identified certain critical aspects of the disease, including axonal degeneration and oxidative stress (Salvadores *et al.,* 2013), dysfunctional mitochondria (Kerr *et al.,* 2017) and imbalanced calcium signalling (Salvadores *et al.,* 2017).

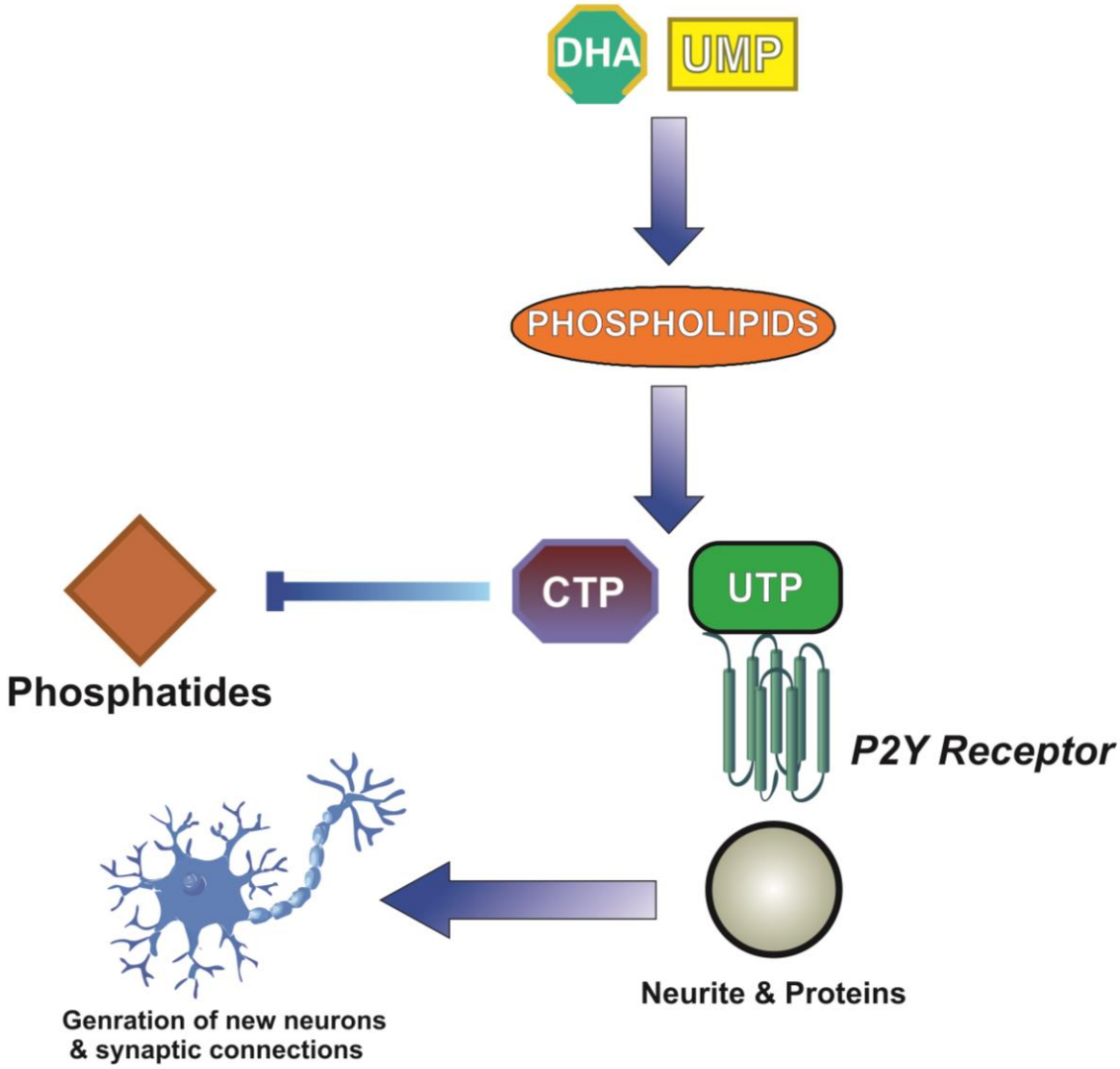

Fig 1.8 A Graphical representation of the action of pyrimidines and their role in improving cognitive functions and memory in the case of neurodegeneration. Uridine monophosphate (UMP), in conjunction with Docosahexaenoic Acid (DHA), stimulates phospholipids synthesis. The phospholipids enhance the formation of other pyrimidine-based nucleotides, Cytidine triphosphate (CTP) and Uridine triphosphate (UTP). CTP is often the limiting factor in the pathway leading to phosphatide formation, although UTP acts via the P2Y receptors that modulate the formation of new proteins and the growth of neurites. This action ultimately leads to novel neurons, slowly impacting overall memory and cognitive functions.

Table 1.3-Defective mutations associated with various genes in pyrimidine metabolism and their relative conditions

Gene with Mutation	Associated Condition
CAD	Delay in Speech, motor development, epileptic encephalopathy, seizures, neurodegenerative disease
Dihyroorotate dehydrogenase	Malformation but no intellectual development issue
UMP Synthase	Intellectual and motor impairment, seizures, 5-FU resistant tumours
CMP Synthase	The proliferation of lymph in the CNS
Thymidine Kinase 2	Mitochondrial DNA depletion, Progressive ophthalmoplegia,
5' Nucleotidases	Developmental delay, ataxia, seizures, speech deficits, splenomegaly, anaemia and mental retardation
Thymidine Phosphorylase	Mitochondrial Neurogastrointestinal Encephalopathy
Dihydropyrimidine dehydrogenase	Motor and Mental Retardation, seizures, epilepsy
Dihydropyrimidinase	Retardation of growth, microcephaly, autism

Most of the drug development against AD is based on the amyloid hypothesis. Drugs reaching up to clinical trials have failed to proceed ahead on more than 99% of the occasion. The primary reason behind this high failure rate in converting the drugs from clinical trials to the market is the delayed diagnosis of pathological conditions (Cummings *et al*., 2014, Salvadores *et al*., 2017). It is now understood that axonal degeneration is one of the critical pathological conditions that occurs pretty early in AD's pathophysiology; it remains a novel drug target. Understanding the mechanism of axonal degeneration could help us target this particular stress for developing drugs (Salvadores *et al*., 2017; Kundu *et al*., 2020). Research has also established that amyloid peptides start accumulating approximately fifteen years before the actual symptoms start to exhibit. This information will also help future researchers develop various methods (invasive/non-invasive) to detect amyloid plaques (Grimaldi, 2018; Kundu *et al*., 2020).

Table 1.4- Purine and Pyrimidine based drug development and their results in some neurological conditions

Disease	Strategy	Target	Outcome
Schizophrenia	Upregulation of adenosine	Enhanced activity of A1 receptors and reduced activity of P2X7 receptors	A decrease in psychosis and better social interaction skills
Autism Spectrum Disorder (ASD)	Suramin and Ketogenic diet	Upregulating A1 receptor activity	Better cognition, social life skills, and language usage
Hypoxanthine Ribosyl Transferase deficiency	S-Adenosylmethionine	No specific target	A decrease in self-injury-prone behaviour
UMP synthase deficiency	Administration of UMP and CMP	No specific target	A decrease in both neurological and non-neurological symptoms
Posphoribosyl pyrophosphate Synthase 1 deficiency	S-Adenosylmethionine	No specific target	Improvement in neurological symptoms
Perinatal white matter injuries	Administration of UDP-Glucose	Activation of GPR17 receptors	Proliferation and generation of newer Oligodendrocytes

Current diagnoses of patients who exhibit a significant decrease in cognitive and associated behavioural changes are taken up for Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans. Positron Emission Tomography (PET) scans are used exclusively in cases where the existence of amyloid β plaques needs to be confirmed. Considering the economic impact, PET scans are costlier than MRI or CT scans and cannot always be used for widespread disease diagnosis. Apart from PET scans, invasive methods like the drawing of cerebrospinal fluid is a method regularly used to test for the presence of amyloid β peptides 1-42 (Dubois *et al*., 2014, Frisoni *et al*., 2017, Grimaldi *et al*., 2018, Kundu *et al*., 2020). The progress of clinical trials is hindered because of the lack of understanding of the pathway of AD's progress. This is mainly due to the lack of reliable biomarkers for detecting the disease's exact stage and over-reliance on invasive methods (Grimaldi *et al*., 2018).

Population-based studies have shown that by 2050 one out of eighty-five people will develop AD, and eightfold of this number will risk showing symptoms. As a result, one of the best

strategies to develop other novel drugs should be targeted as disease-modifying therapies (DMT). Disease-modifying therapies are essential from the perspective that delay in the onset of the disease even by a margin of one year in the coming decade would reduce the overall burden of the total number of patients developing the disease by 2050 by a significant margin (Cummings *et al.* 2018, Kundu *et al.* 2020). Besides developing synthetic drugs, natural compounds are also in the race subjected to developing disease-modifying therapeutic options. The importance and significance of natural drug-based compounds are prevalent from a wide range of ethnobotanical knowledge, which is already reported to affect the improvement of memory and cognitive functions positively. Studies on the Mediterranean diet, rich in vegetables, fruits, cereals, and olive oil, have improved and reduced the chances of developing dementia and dementia-related disorders reported in a cohort study in New York.

Further, diets containing a high amount of polyphenolic compounds, flavonoids, and liquor like wine containing resveratrol also substantially prevent or delay the progress of dementia and the number of associated conditions. Other compounds are now well established and run in various clinical trials like Curcumin and Resveratrol. Other than Curcumin, which is readily available through diets, especially in South-East Asia, Epigallocatechin-3-gallate, myricetin, oleocanthal, and oleuropein are commonly available in green tea is reported to have ameliorating effects and enhance cognitive abilities (Arntzen *et al.* 2010, Eisele *et al.* 2015, Yamada *et al.* 2015). One of the other strategies now well in place is developing potential amyloid fibril formation inhibitors. They are further categorised as (1) agents which could stabilise the native state of the protein, prevention and aggregation (2) remodelling or degradation of already formed amyloid fibrils or insoluble oligomers. Other approaches also worked upon include metal ion chelators, especially iron and copper (Figure 1.9). These two metal ions are essential in forming amyloid β peptides and are now an active drug development target (Hamley 2012; Yang *et al.* 2014, Kundu *et al.* 2020).

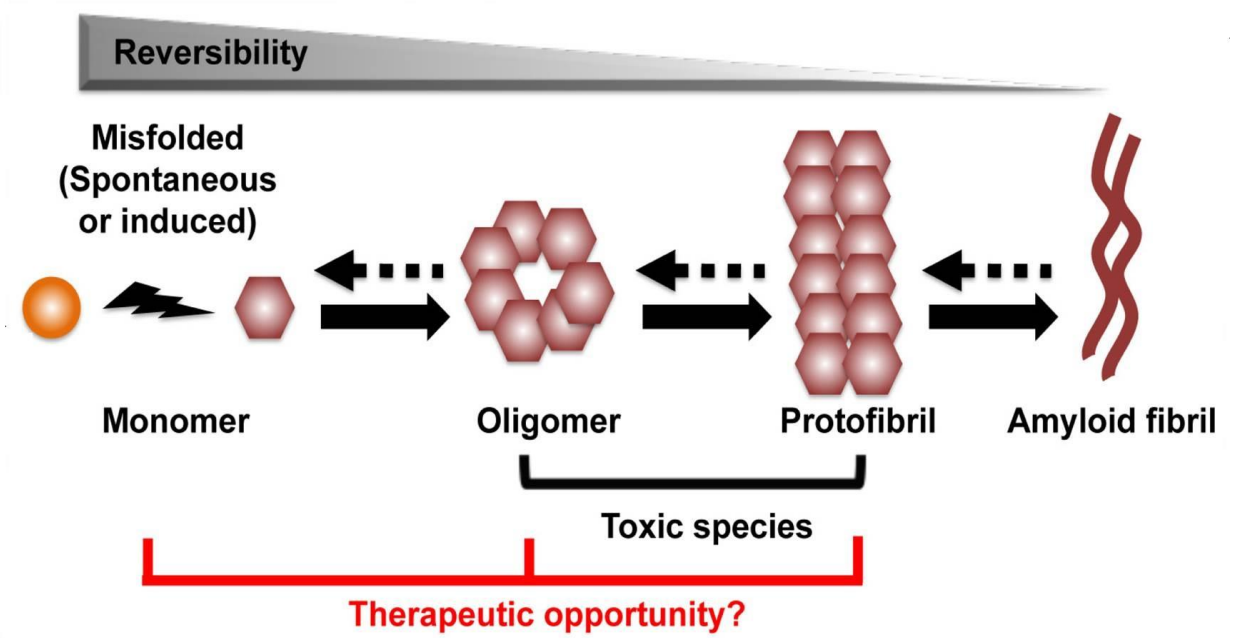

Fig 1.9 A general illustration of the central hypothesis of the present study. The reversibility of the protein aggregated state is much higher in the oligomeric to protofibrillar states. Fully matured amyloid fibrils are highly stable structures and are poor therapeutic targets. Further, oligomeric intermediates are more cytotoxic than amyloid fibrils and are more suited for therapeutic development (*adapted with permission from*: Huang, Liang, Xiaomin Su, and Federoff, J.H. (2013).

1.10- Model Proteins used in this study

Hen Egg White Lysozyme (HEWL)***:*** Hen egg-white lysozyme is one of the most readily available proteins used extensively in studies associated with amyloids and protein aggregation. It is a small protein with 129 amino acids and a molecular weight of 14.3 kDa. It catalyses the beta glycosidic bond breakage between N-acetyl muramic (NAM) acid and N- acetylglucosamine (NAG) in the bacterial cell wall. One of the primary reasons for its usage in this research domain is high homology (60%) with human lysozyme, which is involved in hereditary non-neuropathic systemic amyloidosis (Booth *et al.* 1997; Khan *et al.* 2014). Structurally it belongs to the α helix+β sheet class of mixed proteins (Figure 1.10), out of which nine residues are negatively charged, and 17 are positively charged. Residues 1-36, 87-129 belong to the α- helical domain and 37-86 belong to the β-sheet domain (Eyeles *et al.* 1994; Khan *et al.* 2014).

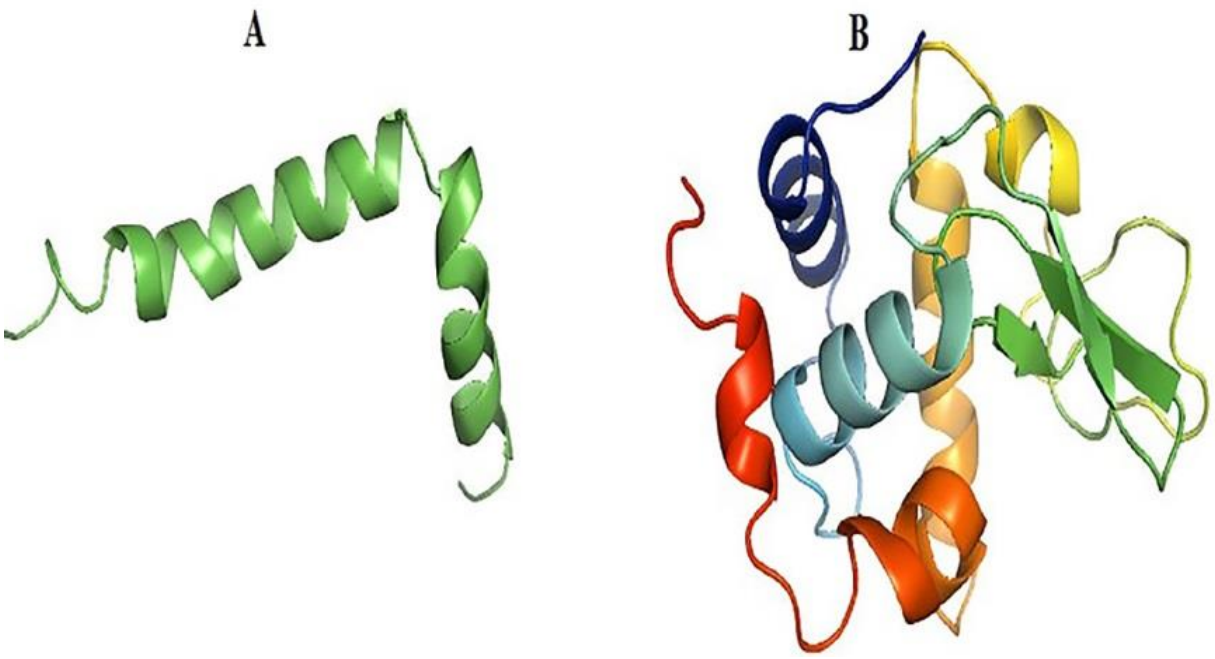

Fig 1.10 Model proteins visualised in Pymol used in this study (A) Amyloid β peptide, PDB ID- 1YT and **(B)** Hen Egg White Lysozyme (HEWL), PDB ID- 1DPX.

1.11- Scope and Objectives of the current study

The lack of sound clarity in understanding the critical phases of protein aggregation is a driving force for low drug discovery in protein misfolding diseases. There is a steep increase in the overall global burden of neurological diseases leading to a high index of mortality and disability among people (Mukherjee *et al.*, 2021). Based on population studies across the globe, reports suggest that AD might be prevalent in every one person among every eighty people by 2050. There is a constant need for disease-modifying therapies (DMT), which are targeted toward delaying the onset and fast progression of the disease (Brookmyer *et al*., 2017, Cummins *et al.*, 2018). Other neurodegenerative conditions like Parkinson's Disease (PD), Huntington's Disease (HD) and Amyotrophic Lateral Sclerosis (ALS) are also projected to be increasing in the next few decades. Multiple therapeutic strategies anticipate discovering essential compounds and pathways that could be utilised for more treatment options. Over the years, research has mainly focused on identifying small molecules as protein aggregation modifiers. Most of the molecules identified belongs to the category of plant metabolites, including various compounds belonging to the category of alkaloids, flavonoids, terpenoids and naturally occurring secondary metabolites. Some of the molecules which have gained prominence over the years include Curcumin, Epigallocatechin-3-gallate, Gallic Acid, *et*

cetera. Although most of these molecules are naturally occurring, they are often required in a high amount of dosage based on results in clinical trials and exhibit bioavailability issues. In the current study, we have rationalised using selected intracellular metabolites as these molecules are endogenous and might have better *in vivo* acceptability. Further, we have also tried to gain some critical insights that these metabolites and the homeostasis of these metabolites would have in protein aggregation mechanisms. With this background, we are presenting the critical objectives of the study.

1. Computational analysis of intracellular metabolites (nitrogenous bases, nucleosides and dNTPs) as potential anti-aggregation agents using model proteins: Hen Egg White Lysozyme (HEWL) and Amyloid β peptide
2. Evaluation of selected intracellular metabolites as anti-amyloid molecules using spectroscopic, microscopic and molecular dynamics and simulation techniques.
3. Effect of selected metabolites on preformed HEWL protein oligomers using spectroscopic, microscopic and protein aggregation kinetics-unravelling mechanistic understanding of their behaviour.
4. Identification of cholinesterase inhibitors using drug repurposing strategy using computational analysis.

The results from this thesis open up some critical areas for a better understanding of protein aggregation, especially in correlation with loss in the homeostasis of intracellular metabolites. The work depicted here also gives us insights into potential mechanisms of small ligands, how they could act as protein aggregation modifiers, and areas where more useful therapeutics could be developed.