Molecular Mechanism of Amyloid Inhibition By Inositol

by

Grace Li

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Biochemistry
University of Toronto

Contents

List of Tables

List of Figures

List of Symbols and Acronyms

$\mathring{A}$	Angstrom
μs	Microsecond
$(GA)_4$	A peptide with Glycine and Alanine repeated four times
Aβ	β-Amyloid
AD	Alzheimer's Disease
ADP	Alanine dipeptide
CNS	Central Nervous System
DSSP	dssp
fs	femtosecond
GROMACS	gromacs
mM	Millimolar
NFT	Neurofibrillary tangles
NpT	npt
ns	nanosecond
NVT	nvt

OPLS-AA/L Potential for liquid simulations-all atom

PME Particle Mesh Ewald

SHAKE shake

STDR simulated tempering distributed replica sampling

TMAO trimethylamine N-oxide

VMD Visual Molecular Dynamics

Chapter 1

Introduction

1.1 Amyloid

Many diseases, most of them neurodegenerative, have been linked with the presence of amyloid: Prion-related diseases, Parkinson's, Huntington's disease, Type II diabetes etc.

1.1.1 Structure and Formation

Finding a treatment for AD and other fatal neurodegenerative diseases motivated many biochemical and biophysical studies of the amyloid state.

1. In vitro, a variety of proteins and peptides, folded or intrinsically disordered, have been shown to be able to aggregate to form amyloid fibrils under certain solution conditions.

2. Amyloid fibrils are formed via a complex aggregation pathway. Initially, monomers aggregate to form oligomers with different morphologies which exists in equilibrium with amyloid fibrils. Some of these oligomers are on-pathway to fibril formation, while others themselves may be end-points of the aggregation pathway. Biochemically, fibrils are protease resistant and are insoluble in the presence of SDS.

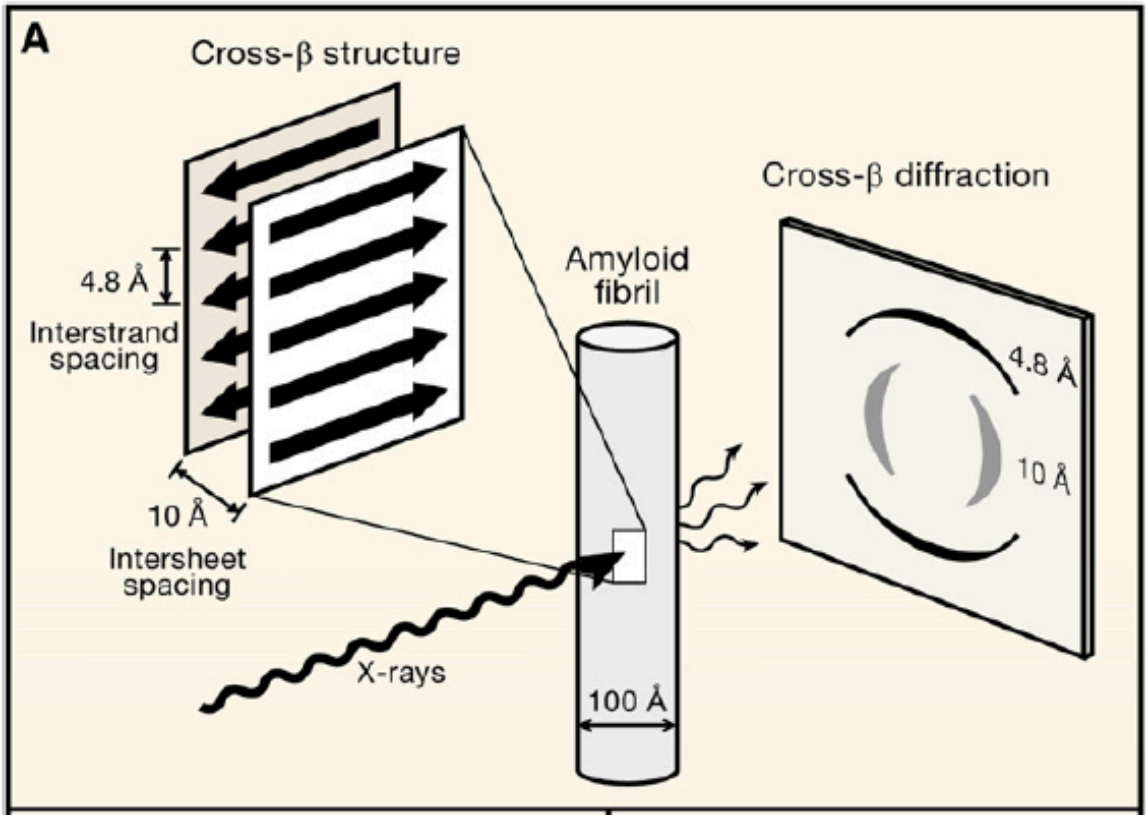

Figure 1.1: This is adapted from Eisenberg, 2012

3. Currently, it is thought that amyloid fibrillar state may be the globally stable folded state for all proteins.

Fibrils

Structure

Early X-ray diffraction studies show that fibrils have a regular structure, which is defined by a characteristic 4.8Å interpeptide, and 10Å intersheet spacing. Biophysicists have adopted this as the definition of the cross-β amyloid fiber structure. (Figure 1.1)

After staining amyloid fibers are visible as long unbranched fibrils under the transmission Electron microscope (TEM). Figure 1.2

In recent years, advances in SSNMR and X-ray crystallography have recently made

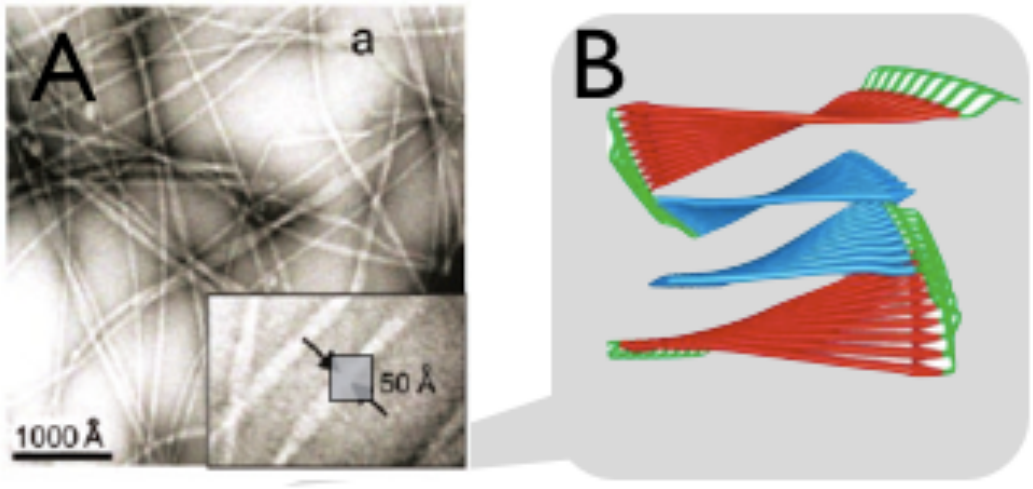

Figure 1.2: A Example EM images of oligomers. Adapted from Bitan G. et al. 2003 and Walsh D. 1999 C TEM image of fibrils D SSNMR model proposed by Tycko et al.

major contributions to our knowledge of the structures of amyloid fibrils.

Both Aβ40 and Aβ42 amyloid fibrils have been studied in detail using solid-state NMR. Figure 1.2

Furthermore, small peptide fragments that have characteristics of amyloid fibrils, which are also amenable to single crystal X-ray diffraction analysis have demonstrated similar type structures from those studied using SSNMR. These structures obtained by X-ray crystallography have been described to have a dry interface with stacked sheets. (Figure 1.3)

"The ubiq- uitous presence of a cross-β structure strongly supports the view that the physicochemical properties of the polypeptide chain are the major determinants of the fibrillar structure in each case. Moreover, several of the proposed structures, despite very different sequences of their component polypeptides, suggest that the core region is composed of two to four sheets that interact closely with each other. An interesting feature of these sheets is that they appear to be much less twisted than ex- pected from the analysis of the short arrays of β-strands that form β-sheets in globular protein structures. This feature was first pro- posed from cryo-EM and has been supported by

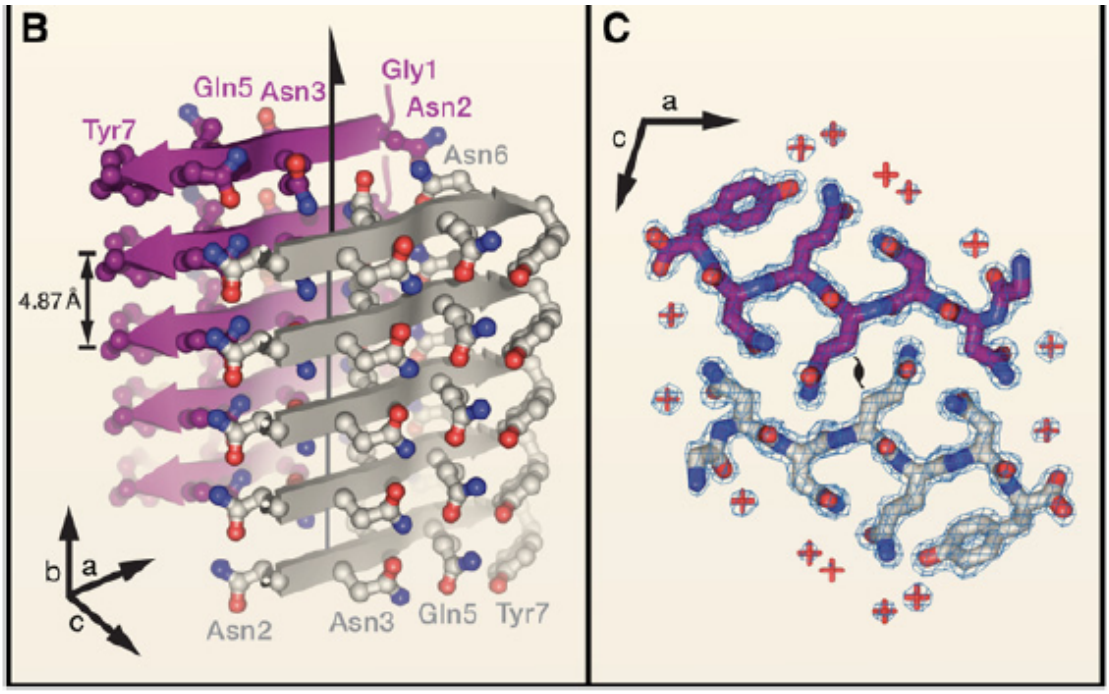

Figure 1.3: This is adapted from Eisenberg, 2012

Fourier transform infrared (FTIR) analy- ses (48, 61)."

Amyloid Fibrils have been found to exhibit polymorphism at the molecular level, but all have similar ultrastructures.

Non-fibrillar oligomers

Due to their structural disorder and their insolubility, structural determination of oligomers have posed challenging experimentally.

1.1.2 Formation

Amyloid fibrils have been observed to form via a two-step nucleation-polymerization process. In the nucleation phase, energetic barriers of aggregation must be overcome to form the initial aggregation nucleus or seed. Following nucleation, free monomers bind to the nucleated aggregates and polymerize into mature fibrils.[2]

1.1.3 Involvement in disease

Known to be involved in different diseases. Neurodegenerative (AD, Parkinsons, Prion), and diabetes, and systematic amyloidosis.

Alzheimer's Disease

One of the most well known amyloid disorders is Alzheimer's disease. AD is the most common cause of dementia in persons of age 65 or older.

With the increasing longevity of our population, AD is approaching epidemic proportions with no cure or preventative therapy available.[1]

MCI progresses to severe memory loss .. what else ...

Neuronal dystrophy is often observed upon examination of deceased AD patients. Furthermore, tissue characterization of dementia brains show the presence of extracellular deposits of neuritic senile plaques and neurofibrillary tangles, which appear as lesions on stained brain tissue under light microscopy. (Figure 1.4)

Although it has been more than one hundred years since Dr. Alois Alzheimer first presented the connection between the presence of neuronal plaques and the clinical symptoms of presenile dementia characteristic of Alzheimer's disease (AD), the exact relationship between the two is still under much contention.

It was not until in the 1980s, the protein Aβ was identified as the largest component of plaques. Aβ aggregates was found to be present in a variety of morphologies in the brain, in fibrillar forms and diffuse forms.

Aβ is produced by intramembrane proteolytic cleavage of the larger amyloid-β precursor protein (APP) by β-secretase, and is produced constitutively as part of the normal cellular metabolism.{Selkoe, 2002 #222} Depending on the position of the cleavage, Aβpeptides of lengths varying from 38 to 43 residues can be produced. However, the peptides spanning residues 1–40 (Aβ40) or 1–42 (Aβ42) are predominantly found AD-

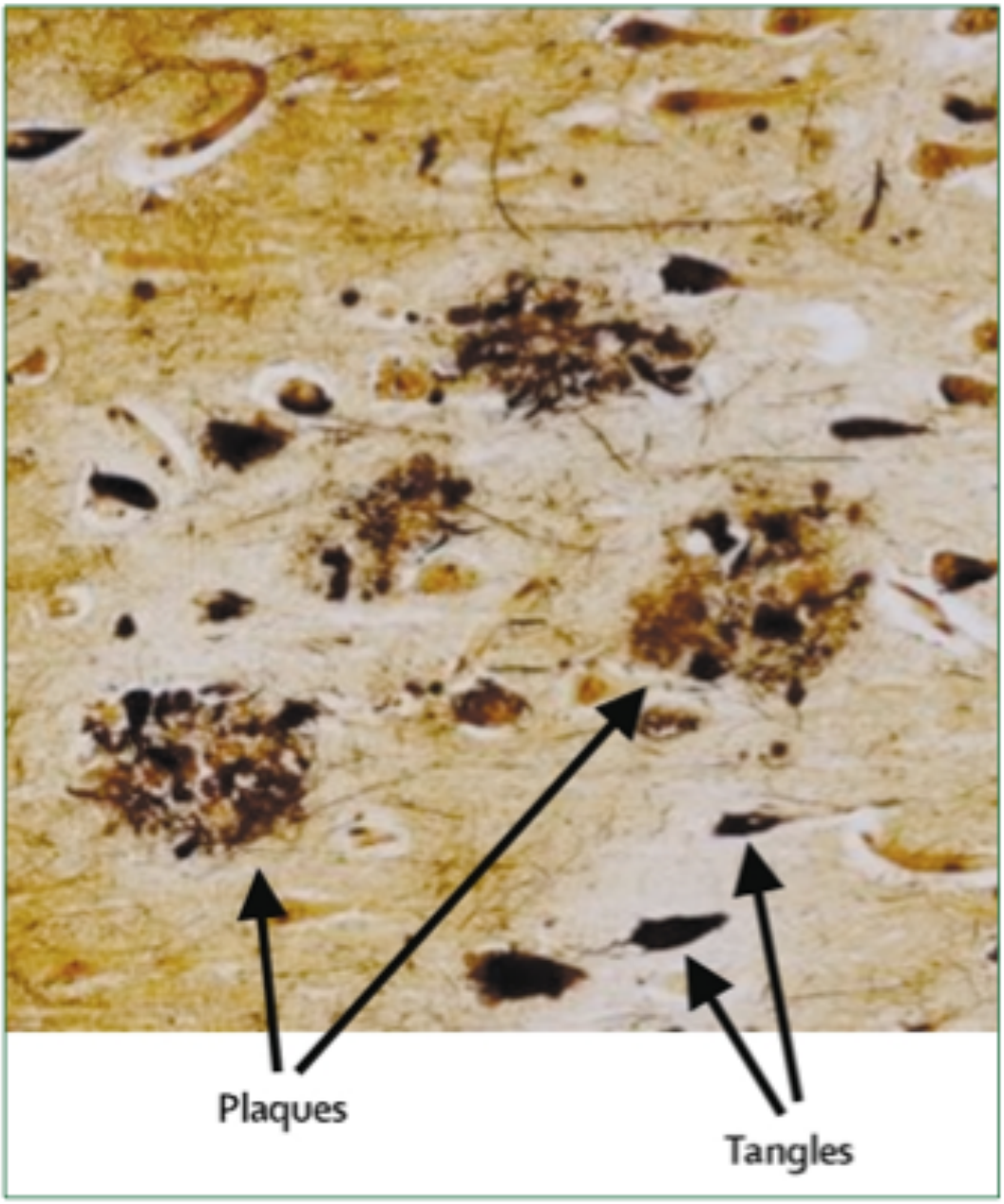

***Figure 1:* Plaques and tangles in the cerebral cortex in Alzheimer's disease**
Plaques are extracellular deposits of Aβ surrounded by dystrophic neurites, reactive astrocytes, and microglia, whereas tangles are intracellular aggregates composed of a hyperphosphorylated form of the microtubule-associated protein tau.

Figure 1.4: This is adapted from Blennow, 2006

associated plaques.

Abeta42 is thought to be more involved in AD than Abeta40. They oligomerize through different pathways REF. Abeta42 aggregates much faster than Abeta40.

The presence of amyloid plaque deposits in brains of deceased dementia patients led to the formulation of the long-standing amyloid hypothesis, which posits that amyloid aggregates initiates the pathogenesis of AD and that the other pathological symptoms such as neurofibrillary tangles are secondary.

Although both plaques and NFTs appear, plaques are thought to more likely cause AD than NFT. NFTs involve the aggregation of tau protein. Several lines of evidence suggest that NFTs plays a secondary role to Abeta in the pathogenesis of AD. Knock out mouse models ... mice do not develop AD, and instead develop tau pathologies NFTs are NFTs have also been shown to be affected by Aβ production.

AD is difficult to diagnose. It is often not apparent that someone has AD until they exhibit symptoms severe enough to interfere with daily life or occupation. Although plaques are often visible in the dementia patients, plaque load does not appear to correlate well with the progression of disease. In some individuals without dementia symptoms may have as much plaque as another with severe AD. Synaptic loss can be used as a measure of disease progression. It has been found in recent years that synaptic loss correlates with the concentration of soluble Aβ oligomers in the brain.

Currently there is a lack of treatment which targets the underlying disease. Most approved treatments today for AD are only for mitigating cognitive symptoms.

Much of the focus on drug development for Alzheimer's has been on preventing amyloid aggregation and decreasing amyloid production. We will discuss this in more detail in later section XXX.

Other disorders

In addition to AD, other neurodegenerative diseases have been shown to involve the presence of amyloid. Parkinson's disease, Huntington's, Prion disorders (Mad cow). These diseases and their pathology are reviewed elsewhere and are beyond the scope of the thesis.

Toxicity

- Multiple lines of research have identified oligomers as a likely causative agent for neuronal cell death. By contrast, the monomeric and fibril forms are thought to be less toxic than oligomers. It is hypothesized that soluble oligomers may cause toxicity by perturbing the integrity of cellular membranes through binding and disrupting the lipid bilayer (perhaps by making them ion permeable).[3]
- Include a paragraph about amyloid formation and lipid membranes (?)

1.2 Amyloid Inhibition by small molecules: A promising method of treatment for amyloid disorders

Amyloids are attractive drug targets. Small molecules which targets amyloids may be an effective method of treatment for amyloid disorders because they have the potential to be able to treat the underlying disease. Through in vitro screening, many small molecules have been found to effect the amyloid aggregation pathway. Some were demonstrated to inhibit amyloid fibrils, where as others were shown to arrest or reduce oligomer formation.

Pharmacological perspective of the challenge of developing an Alzheimer's drug. In order to effectively treat Alzheimer's and other neurodegenerative diseases, small molecule drug candidates must pass the blood brain barrier at sufficient concentrations for inhibition. This is difficult to achieve.

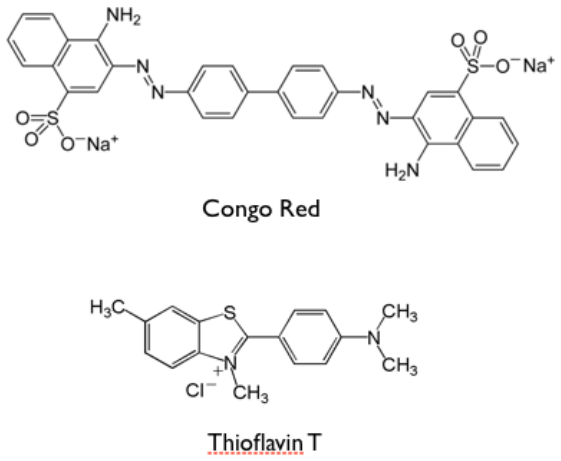

Figure 1.5: Amyloid binding dyes Congo Red and Thioflavin T

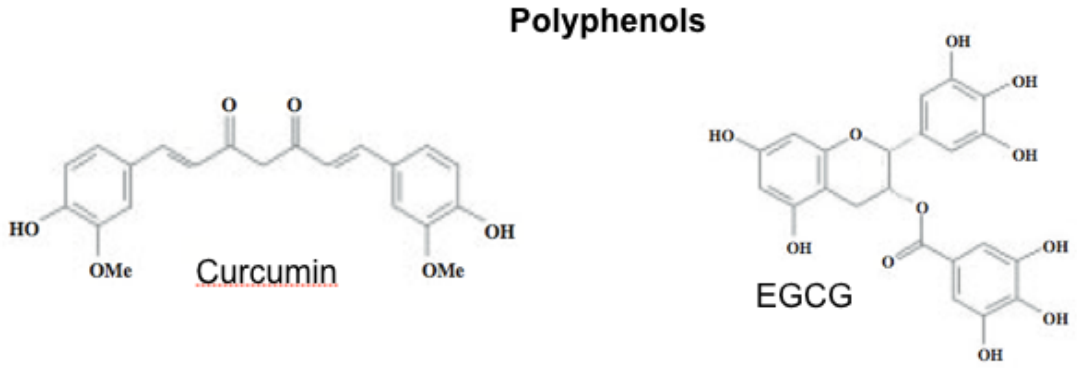

Figure 1.6: Polyphenols

In vitro screening has led to the discovery of a large number of small-molecules which were found to affect the amyloid aggregation pathway. Many of these small molecules are thought to act by directly binding to amyloidogenic peptides and aggregates.

1.2.1 Dye molecules

Thioflavin T and Congo red are dye molecules used to identify the presence of amyloid. Both molecules bind to mature amyloid fibrils and have been shown to affect fibril formation.(Fig. 1.5)

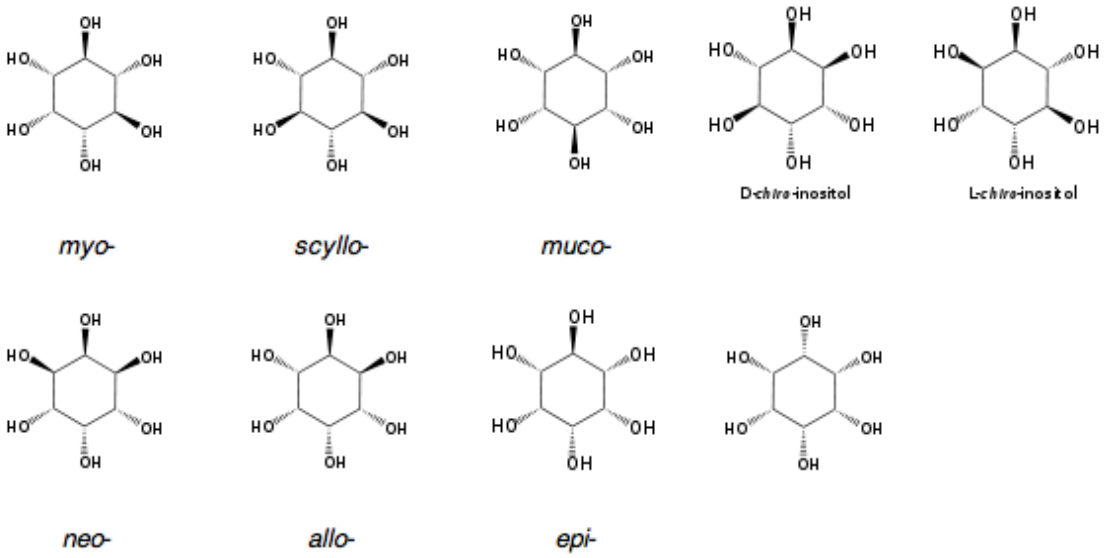

Figure 1.7: Inositol stereoisomers

1.2.2 Polyphenols

Polyphenols, is a large group of natural and synthetic molecules. (−)-epigallocatechin-3-gallate, curcumin, and a polyphenolic grape seed extract, known for their anti-oxidant properties, were recently discovered to be capable of affecting amyloid formation.(Fig. 1.6)

1.2.3 Inositol

Inositol stereoisomers.(Fig. 1.7) Role of inositol in the human body. *myo*-inositol. Inositol is not toxic to the human body. Inositol is used in signaling pathways. Furthermore, scyllo-Inositol is able to cross the blood brain barrier. It has high bioavailability. Because it is not broken down in the gut, it can be taken orally.

Inositol stereoisomers are found in various human tissues. It is particularly found at high concentrations in neuronal tissues. Inositol is synthesized inside the body ... or can be obtained via nutrition. Myo- is present in certain grains, grape fruit, but scyllo- is only found in small quantities in food sources.

1.2.4 Commonalities between small molecule inhibitors

Small molecule inhibitors share common chemical features and groups. They are typically planar in geometry, have many aromatic rings, and polar functional groups (hydroxyl groups) around the edge of these aromatic rings.

1.2.5 Molecular mechanisms of amyloid inhibition by small molecules

Some small molecules inhibit fibril formation, where as others may prevent oligomerization, but not fibrillation. A high concentration is often required to observe activity (micromolar to millimolar), which suggests that they may be non-specific inhibitors. EGCG, one such polyphenol, is known to have the lowest IC50.

Molecular mechanism of binding of dye molecules. Thought to bind flat on on the surface grooves of amyloid fibrils where they interact with hydrophobic groups exposed at the surface.

1.3 Analogy to Sugar-protein binding

1.3.1 Sugar Binding modes

1.4 Protein-ligand interactions

1.4.1 Forces involved in binding

- Protein-ligand non-covalent interactions that are important for ligand binding and recognition
  - Electrostatic interactions. Polar (hydrogen bonding) and charge-charge interactions

  - Nonpolar (hydrophobic) interactions
    * Van der Waals

1.4.2 Binding equilibria

- Enzyme and its putative ligand typically bind specifically (high affinity binding). We want to optimize binding specificity to increase the efficacy of the putative drug, and decrease adverse side effects (toxicity) in the human body.

- The dissociation constant, K_d, is a measure of the affinity of a ligand for its binding site on the host protein. Pharmacologically, it can be interpreted as the concentration at which 50% of the drug is bound to the protein. In experimental studies, K_d is often used to quantitatively screen for potential drug candidates.

- Binding equilibrium

$$[Protein \cdot Inositol] \rightleftharpoons [Protein] + [Inositol] \tag{1.1}$$

- The binding free energy of a ligand to a protein is directly related to its dissociation constant, K_d, the equilibrium constant of the above reaction

$$K_d = f_{ub} \frac{[Protein]\,[Inositol]}{[Protein \cdot Inositol]}, \tag{1.2}$$

- Experimental techniques for estimating K_d
  - What experimental techniques are used to estimate binding affinity? (May need to study up on this)
  - Isothermal titration calorimetry (ITC) is a technique which can be used to measure energetics of ligand binding to peptides.

1.5 Thesis objectives and rationale

1.5.1 Challenges of amyloid inhibition

- The protein-ligand binding model developed to understand enzyme inhibition cannot be directly applied to understand the molecular mechanism of amyloid inhibition by small molecules.
  - Amyloid inhibitors are found to be very weak binders. How do non-specific inhibitors act as a drug? And how do we approach this with MD simulations?
  - Because the Aβ amyloid aggregate pathway encompasses a variety of species, some of which has no folded structure, a single conformation cannot be assumed for binding. Furthermore, structural information of amyloidogenic species lags behind those of enzymes, which tends to be globular proteins amenable for X-ray crystallography. This means that the putative binding sites are not known.
  - The structural disorder of the peptides involved poses a challenge for obtaining converged properties from MD simulations.
- Aβ peptides are completely disordered. We also do not know what the binding site looks like, where it is located on these structures.
- To date, few studies have attempted to provide statistically meaningful results pertaining to general mechanisms of protein self-aggregation and amyloid formation. Furthermore, despite the abundance of MD studies of Aβ, few studies have systematically examined the mechanism of action of small molecule inhibitors of amyloids
- In AD, there is the added challenge of the drug being able to cross the brain barrier, while remaining non-neurotoxic. What kind of drugs cross the BBB? Typically

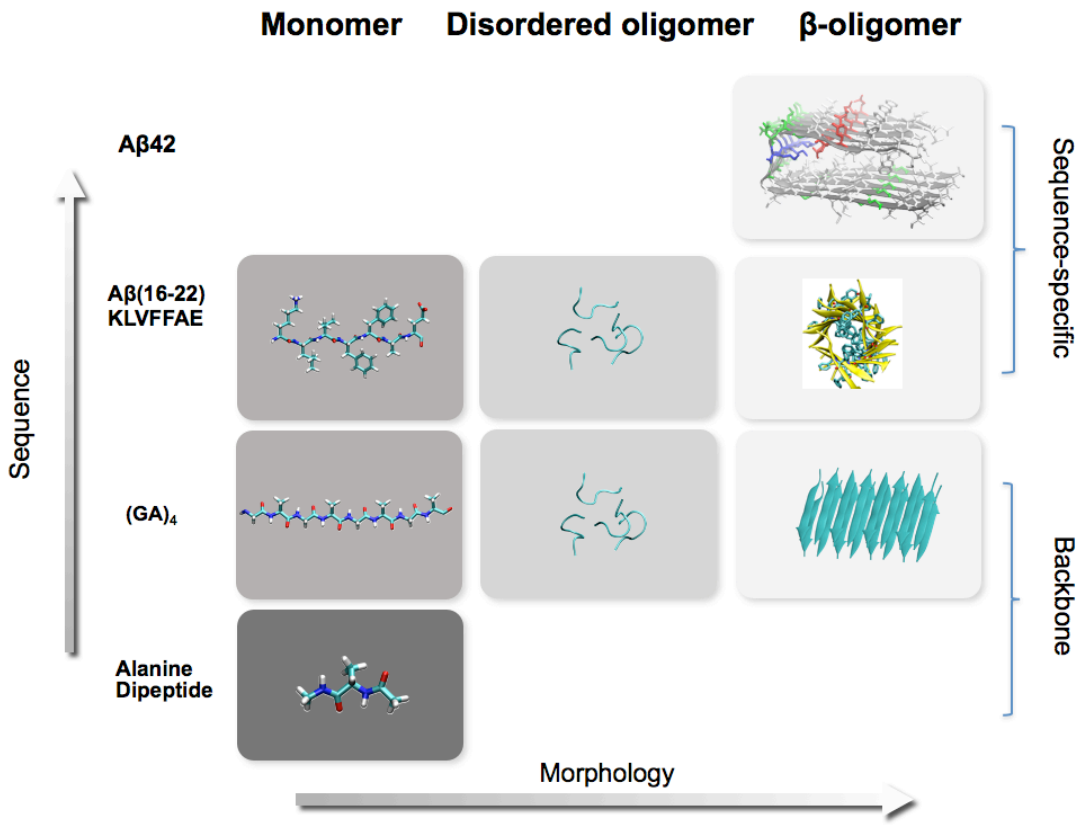

Figure 1.8: Shows the progression from small, model systems to larger and structurally more complex systems involving the full-length Aβ42 peptide.

hydrophobic drugs.

1.5.2 Study Design and Rationale

- Here describe in detail how I designed my study to circumvent the challenges presented by the amyloid inhibition problem, and the limitations of MD simulations. At this point, clearly explain and discuss my study design and rationale. (Fig. 1.8)

- Beginning with the simplest model systems for an amyloidogenic peptide, the alanine dipeptide, we systematically examine binding of inositol with systems of both increasing sequence and structural complexity.

- We exploit conventional MD simulation techniques because simulation approaches

used for understanding enzyme-ligand binding is not applicable.

- Instead, we use conventional MD simulations and repeats of independent simulations to determine the binding modes, and binding equilibria of inositol with amyloidogenic peptides and aggregates of Aβ.

1.6 Thesis Organization

The first chapter introduces the thesis in the context of the field. The second chapter introduces the main methods used in the work in the thesis. Chapters 3, 4, and 5 are the results of simulations of inositol with amyloid like peptides and aggregates. Chapter 6 shows work of the general applicability of our methods developed throughout this thesis to MD simulations to understand protein - carbohydrate binding. Chapter 7 provides discussion, suggestions for future work, and perspectives.

Bibliography

[1] Kaj Blennow, Mony J De Leon, and Henrik Zetterberg. Alzheimer's disease. *Seminar*, 368:387–403, 2006.

[2] Regina M Murphy. Peptide aggregation in neurodegenerative disease. *Annu Rev Biomed Eng*, 4:155–174, 2002.

[3] Dominic M Walsh and Dennis J Selkoe. A beta oligomers - a decade of discovery. *J Neurochem*, 101(5):1172–1184, 2007.

Chapter 2

Methods

2.1 Molecular Dynamics Simulations

MD simulations employ an empirical mathematical function to describe the atomic interactions in a molecular system, and together with classical laws of Newtonian mechanics, atomic trajectories of motion are generated.

Thermodynamic and kinetic properties can then be extracted as time averages from these trajectories and used to make a number of predictions that are often experimentally challenging to observe or measure.

MD simulation studies have been useful in studying many existing fundamental problems of biology and biochemistry, including protein dynamics and function, protein folding, biomolecular self-aggregation, and protein-ligand binding.

Under the Born-Oppenheimer approximation, electronic and nuclear motions are uncoupled, and can therefore be treated separately. In MD atomic nuclei are treated as classical particles, where the movement of electrons are not accounted for. Instead, the effects of electronic motion is implicitly accounted for via the use of potential energy functions.

The force on each atom in the simulation is the spatial derivative of the potential

energy of the system,

$$F = -\nabla U \tag{2.1}$$

where U is the potential energy of the system. In MD simulations, U is estimated using a molecular mechanics force field, a set of parameters which describes the interaction energies of the atoms in the system.

The acceleration a of each atom is then given by Newton's second law of motion,

$$a = \frac{F}{m} \tag{2.2}$$

where m is the mass forces on each atom.

Dynamics of the next position of the atom at time $t + \delta t$ is then

$$x_i(t + \delta t) = x_i(t) + v_i(t)\delta t + \frac{a_i(t)\delta t^2}{2} \tag{2.3}$$

Once all of the new positions for all atoms are predicted, updated interatomic forces based on the coordinates in the newly generated frame are calculated, and the entire process is repeated.

For numerical stability reasons, a small integration step or timestep is used in the calculations. Typically 2 femtosecond timestep, twice the time period for the fastest vibrational motion (for bonds involving hydrogen), is used in MD simulations of biomolecular systems.

Application of an empirical force field is used in MD to approximate atomic interactions in the system. A force field typically has many parameters which need to be calculated. One approach to do this is to fit to quantum mechanical calculations. Often, force fields are iteratively improved by predicting experimentally observable quantities for small compounds, and adjusting the fit based on comparisons of these computationally predicted quantities with experimental measurements.

There many different force fields (AMBER, Gromos etc), each differing slightly in the potential energy function and its parameterization. In this thesis, we have chosen to use the OPLS-AA/L force field for all of our simulations.

The general form of the force field potential energy function is

$$\begin{aligned} E = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{dihedrals} k_\chi(1 + cos(n\chi - \delta)) + \sum_{impropers} k_\gamma(\phi - \phi_0)^2 \\ & + \sum_{nonbonded} \frac{q_1 q_2}{er} \\ & + \sum_{nonbonded} \epsilon[(\frac{r_{min}}{r})^{12} - 2(\frac{r_{min}}{r})^6] \end{aligned} \tag{2.4}$$

2.1.1 Setting up a MD simulation: practical aspects

The following steps are often used to setup and start a MD simulation system of a protein. First, a pre-determined structure, typically a coordinate structure from X-ray crystallography or NMR, or homology-modelling data. Then a force field and solvent is chosen.

2.2 Challenges and limitations of MD

- MD is computationally challenging because of limitations in length and time scales.
  - Length scale. Large systems are too complex to obtain statistics and quantitative predictions.
  - Time scale. Relevant biochemical reactions such as protein folding happens on the time scale of milliseconds, hours, and days. Currently with MD simulations, we are routinely able to approach the microsecond time scale, massive

computing power is still insufficient to observe phenomena on the millisecond timescale.

- Obtaining convergence. Explain why this is difficult, in particular for systems with disordered peptides.

- Limitations in the accuracy of current force fields.

2.3 Application of MD in structure-based drug discovery

- A broad application of simulations of proteins is to computationally design drugs and combining that with structure-based drug design. In recent years structure-based computer modeling of protein-ligand interactions have become a core component of modern drug discovery.

- Current drug discovery platform. Typically, the first step in drug discovery is to identifying a target, a putative binding site. Then, solve the X-ray crystal structure of the target.

- Ligands which may act as potential drugs typically have high binding affinity to the binding site. The goal is to find high specificity inhibitor of a protein (usually an enzyme). The binding free energy is an important quantity which can be used to evaluate how well a ligand binds. One method of estimating binding affinity is by using computational docking methods, where the binding affinity is typically estimated without taking into account of protein dynamics. Although docking is computationally less expensive, it is inaccurate for identifying true drug candidates because binding often involves crucial changes in protein conformation.

- With computer hardware becoming faster and cheaper, MD simulation and mod-

eling can be used to rapidly prototype experimental ideas – for example, one can perform computational alchemy, that is, "mutate" residues to test various hypotheses. Furthermore, simulations may be used to determine whether a chemical change will produce a more potent drug candidate.

- Currently state of the art computational binding studies take into the account of change in protein conformation. MD simulations is an effective method, where the protein and drug is allowed to relax and freely move about in the system.

- However, in the case of understanding a specific binding reaction (eg. when developing an enzyme inhibitor), the ability to observe the relevant binding events is a low probability event on the timescale achievable in our simulations. Therefore, it is impractical in this case to solely apply brute-force sampling techniques to determine binding free energies.

  - Methods used to determine binding free energies using simulations:

    * Thermodynamic perturbation[1]

      · thermodynamic integration

      · free energy perturbation

2.4 Review of MD studies of amyloid inhibition by small molecules

- In recent years, molecular dynamics simulations have been intensively used to investigate the molecular basis of the structure and stability of amyloid fibrils.

- MD simulations of Congo red binding have been done with the protofibril-like crystal structure composed of the segment GNNQQNY.{Wu, 2007 #621}

- A recent simulation study of an N-methylated peptide with Aβ16-22 models of amyloid aggregates has provided insight into the possible mechanism of action of peptide inhibitors of amyloid formation.{Soto, 2007 #597} This peptide inhibitor was shown to preferentially bind monomers to form dimers, possibly acting to inhibit fibril formation by sequestering monomers. However, peptide-based inhibitors have poor pharmacological profiles as they are actively broken down by proteases in the stomach and are difficult to transport across the blood-brain barrier. In addition, these peptide inhibitors specifically target Aβ and thus do not have the potential to treat multiple amyloid diseases.

Bibliography

[1] Michael K Gilson and Huan-Xiang Zhou. Calculation of protein-ligand binding affinities. *Annual review of biophysics and biomolecular structure*, 36:21–42, 2007.