Application of computer simulation to the study of neurodegenerative diseases

by

Lei Lou

(Under the Direction of Zhong-Ru Xie)

Abstract

Neurodegenerative diseases threaten human life expectancy, and their pathology remains unclear. These diseases share in-common features that distinctive secondary structural changes are often recognizable along with protein aggregations, which can be used to define whether a lesion has occurred. However, it is difficult to observe the dynamic process by traditional biological methods. Abnormal secondary structural changes are often a recognizable feature of these diseases and are usually accompanied by more pronounced signs of aggregation. In this dissertation, computational approaches are used to model the denaturation of crucial proteins in neurodegenerative diseases and examine different hypotheses with their supporting evidence. Evidence about certain neurodegenerative diseases was validated by representing the structural changes of those features. Alzheimer's disease and human prion diseases are the main subjects of study, and the initial stages of their protein misfolding are simulated and revealed.

For Alzheimer's disease, we identified the possibility of endogenous factors that induce misfolding by building a model only containing Amyloid beta (Aβ) peptides and revealing their misfolding process. The Molecular Dynamic

(MD) simulations of the Aβ dimers show a misfolding transmission process between a normal Aβ and a misfolded Aβ; misfolding usually starts with I32-L34. Miquelianin was virtually screened out from the drug database as an inhibitor candidate that can prohibit the formation of the beta sheets. Heparin was identified as a stabilizer for maintaining the parallel polymer structure. With heparin in the system, the parallel structure of Aβ will not be disrupted in the MD simulation, which may cause fibrillation Aβ. The electrical charge on the Aβ surface in different pH reveals the active site of Aβ to receive other ligands. Moreover, an exogenous factor that triggers misfolding was also discovered. Herpes Simplex Virus type 1(HSV-1) glycoprotein B was identified as an etiological agent. Structure prediction algorithms were applied for human prion disease to generate the N- terminus of the whole protein sequence. The formation of beta sheets 129M-130L and 162Y-163Y was regarded as a sign of the start of a misfolding. The predicted structural models reveal that the differences in spatial location between the N-terminus and the primary domain of prion protein may determine whether the misfolding starts. A new misfolding process of prion is proposed based on the simulation results and previous study.

INDEX WORDS: [Neurodegenerative disease, Alzheimer's disease, prion, computer-assisted, simulation,]

Application of computer simulation to the study of neurodegenerative diseases

by

Lei Lou

B.S., Chongqing University of Technology, China, 2017

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2022

Application of computer simulation to the study of neurodegenerative diseases

by

Lei Lou

Major Professor: w Zhong-ru Xie

Committee: Bingqian Xu
Changwei Li
Houjian Cai
Leidong Mao

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2022

Acknowledgments

I express my immense gratitude to Dr. Zhong-Ru Xie, who provided me with exceptional guidance on genomics research. Words cannot express my gratitude to my him for his invaluable patience and feedback. His high standard and cautious attitude toward science have significantly impacted my path to becoming a bioinformatics scientist. I could not have undertaken this journey without my defense committee, who generously provided knowledge and expertise.

I am also grateful to my classmates and cohort members, especially my office mates, for their editorial help, late-night feedback sessions, and moral support. Thanks are also due to the university librarians, research assistants, and research participants who influenced and inspired me. I also want to thank members of Xie's lab that I have worked with over the past years: Dr. Yifei Wu and Godfrey Hendrix. They have been my great colleagues and collaborators.

Lastly, I would be remiss in not mentioning my family, especially my parents, and my girlfriend. Their belief in me has kept my spirits and motivation high during this process. I would also like to thank my cat for all the entertainment and emotional support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter will cover the backgrounds, basic knowledge, and technical resources.

1.1 Computational Drug Design

Computational drug design integrates innovative computer technology with the conventional drug industry, and the main need to implement it is to make the whole drug discovery process more efficient and cost-friendly. (Prieto-Martínez et al., 2019; Young, 2009)

1.1.1 Conventional Drug design

Conventional drug design pipelines usually include target identification, drug screening, and drug candidate optimization. They highly rely on the specific biological clue of a target or certain existing drugs, most of which can be origins from serendipitous discoveries. Identification and screening jobs for validating a promising drug candidate usually take decades and cost billions of dollars. In recent times, drug discovery has included many steady steps in the laboratory: identifying and validating targets, generating assays to find lead compounds,

and optimizing lead compounds to increase affinity and efficacy and reduce potential side effects. Although drug development cycles are significantly shorter, new diseases frequently emerge with increased population mobility and environmental changes. Human public health is still facing severe challenges. In this context, the need to find an efficient method to find novel drugs is an urgent need to be addressed. (Alonso et al., 2006; Hajduk & Greer, 2007; Panigrahi et al., 2012)

Developing drugs can take much longer time than we think. On average, it takes about 14 years to get a drug candidate from the lab to the market, costs more than 1 billion dollars, and has a low success rate. A successful drug will pass through all five phases: drug discovery, preclinical studies, clinical trials, FDA approval, and post-marketing surveillance. Success statistics are acquired at several stages of drug development; a new report states that 13.8 percent of drug candidates entering the first phase of clinical trials eventually receive approval from the FDA, the first test of its kind in humans. After gaining approval, companies have a period of exclusivity ranging from six months to seven years to earn back the enormous costs of drug development before competitors release generic versions. For this reason, the economic drive associated with pharmaceutical companies is necessary so that they can continue to produce life-saving drugs. (G. Lee et al., 2019)

A target is a macro-molecular object (such as a protein) that performs a biolog-

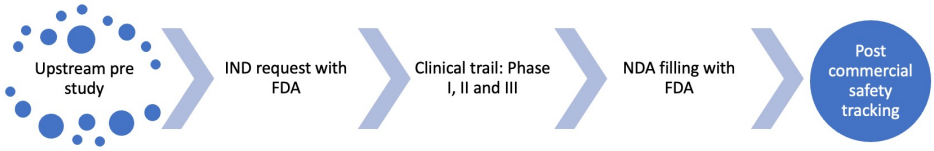

Figure 1.1: The pipeline for conventional drug design

ical function. It is also called a receptor. A drug target has a specific effect on a

disease. Drugs are molecules or ligands that bind to receptors or drug targets. A good drug target should be safe and meet clinical and commercial requirements. Drug target identification begins with identifying the function of a possible therapeutic target, either a protein or a gene that dedicates the structure of the protein (gene targets that are related to specialized drugs and will not be discussed here). Drugs to be developed will be used to attack or modify target proteins related to specific diseases or health matters that require medical intervention. Experimental models and assays are developed during target validation to screen and assess pharmacological associations with phenotypes of interest. Validation can be done using tool compounds (chemical genetics) or genetic methods. These efforts confirm the link to the disease of interest but also that the modulation of the target is safe.

1.1.2 Drug Screening Methods

Since the invention of computers, most traditional industries have been highly integrated with computers to achieve industrial upgrading, and the pharmaceutical industry is no exception. People built computational algorithms to store molecule information in a virtual world and tried to utilize it for virtual testing. The two best experimental screening methods are high-throughput (HTS) and focused or knowledge-based screening. (Hajduk & Greer, 2007) HTS uses automated robotics to quickly perform millions of assays to match the drug-like properties, with extensive libraries of compounds assuming no prior knowledge of the compounds. Focused screening involves narrowing compounds down to a smaller subset that may have particular prior-known activity with the drug target through literature or patent precedents.

1.1.3 Drug Re-purposing

For existing drugs, drug re-purposing can be done by computational simulations to have a more comprehensive knowledge of them, which is important for building a comprehensive understanding. (Oprea et al., 2011; Pushpakom et al., 2018) This thesis focuses on computer-assisted drug re-purposing, which is fundamentally a re-scan of the receptor for the drug itself. Drug re-purposing can be divided into two categories: the first is to target the same receptor protein, not to change the essential drug-receptor model, but to look for other diseases or pathologies affected by this receptor protein, such as Minoxidil, which was converted from an anti-hypertensive vasodilator anti-hair loss drug; the second is to look for a completely new target protein for the drug, and after determining that it has a more desirable binding capacity, to relate it to the biological mechanism in which it may be involved. (Rudrapal et al., 2020) Computer-assisted drug re-purposing is mainly the second approach mentioned above, in which the candidates are obtained in a relatively short time by computerized characterization in a protein database.

Although drug repurposing is not new, it has become an important strategy of the pharmaceutical industry and can occur at all stages of drug development. Many pharmaceutical companies have recently developed new drugs by applying drug repurposing strategies. Repurposing drugs to treat common and rare diseases offers the advantage of using de-risked compounds. Repurposing drugs for new indications is a way to find treatments for rare diseases that offer advantages over drug development due to cost-effectiveness and time efficiency. Drugs developed for one indication may have expanded applications in that particular disease state. ("Drug Repurposing: Advantages and Key Approaches | Technology Networks", n.d.)

Table 1.1: Differences between Biological and Computational Drug Repurposing

Activity-based approach	**In silico-based approach**
Experimental (in vitro and in vivo) screening	Computational (virtual) screening
Target-based and cell/organism-based screening assay	Protein target-based screening
Requires no structural information of target proteins and drug-induced cell/disease phenotypic information	Requires structural information of target proteins and drug-induced cell/disease phenotypic information
Time and labor consuming	Time and labor efficient
Lower rate of false positive hits during the screening	Higher rate of false positive hits during the screening

Table 1.2: Drug Re-purposing Resources

Methodology	Type of method, category	Method/specific approach	Example(s)
Drug-oriented			
Phenotypic screening	Blinded/ Target-based, Screening	In vitro and in vivo HTS/HCS screening	Sildenafil (erectile dysfunction), rituximab (breast cancer)
Target 3D structure, chemical structure, information of drugs and ligands	Target-based, Cheminformatics	In silico screening, ligand-based screening and molecular docking, fragment-based screening	Fluorouracil (lung cancer), etoposide (bladder cancer)
Drug-target information, chemical structure, information of targets and drugs	Knowledge-based, Bioinformatics/ Chem-informatics	Drug–target prediction	Simvastatin, ketoconazole (breast cancer)
Clinical trial information and adverse effects	Knowledge-based, Bioinformatics	Drug similarity studies	—
FDA approval labels	Knowledge-based, Bioinformatics	Drug similarity studies	—
Disease-oriented			
Available Pathway information	Knowledge-based, Bioinformatics	Discovery of disease mechanism and address of key targets	Vismodegib (skin cancer)
Disease omics/ genetics data	Signature-based Bioinformatics	Studying gene signatures/ genomics to identify key targets	—
Disease omics data, available pathway information, and protein interaction network	Pathway or network- based, Network biology	Analysis of disease-specific pathways and networks to identify key targets	Sunitinib, dasatinib (breast cancer, brain tumor)
Therapy-oriented			
Drug omics data	Signature- based or Signature- and network- based, Bioinformatics and/or Network biology	Studying gene signatures	Sirolimus (acute lymphoblastic leukemia), Fasudil (neurodegenerative disorders)
Disease omics and drug omics data	Signature based, Bioinformatics	Similarities between drugs and diseases	Cimetidine (lung cancer), topiramate (inflammatory bowel disease)
Drug omics data, disease pathway and protein interaction network	Targeted- mechanism based, Network biology and Systems biology	Elucidating targeted pathways	Daunorubicin, clomifene (breast cancer)

1.1.4 Bioinmormatics

Researchers often find drug targets from academic studies, scientific literature, or bioinformatics data mining. The advantage of data mining is that it enables selection and prioritization when looking for potential disease targets. (Blundell et al., 2006) With the advent of artificial intelligence (AI) methods, bioinformatics data mining has advanced considerably. After the target has been identified and selected, the molecular target needs to demonstrate a functional role that has a therapeutic effect on the onset or progression of the disease. (Mandal et al., 2009; Ndagi et al., 2020) Bioinformatics approaches for drug discovery include anything related to the biological function of a potential drug candidate, including sequence-based features, interactions with body structures (metabolites, proteins, cells, tissues, etc.), pathway perturbations, and toxicity. Multiomics and high-throughput sequencing are also major topics of bioinformatics. Most bioinformatics sub-disciplines can be applied to the drug discovery process. (Wishart et al., 2018; Wollacott et al., 2019)

One of the most direct ways bioinformatics can help drug development is by predicting the structure of proteins, shifting from a sequence to a three-dimensional structure. By incorporating three-dimensional structural information, the accuracy is further improved compared to traditional predictions of binding affinity based solely on the sequence. The main advantages of building protein structure models include overcoming limitations for specific proteins that can not be acquired by conventional methods; making the precise prediction for the unstable domains, including the N-terminus; and saving time and money with reduced cost.

Based on the available resources, bioinformatics is not necessary for drug development, but it is a great impetus because of the gradual improvement of genetic databases, which makes gene-level studies different, and the emergence of next-generation sequencing (NGS), which has led to significant progress in

genome-wide studies. By using NGS, the analysis of epigenetic factors such as genome-wide DNA methylation and DNA-protein interactions will also be achievable. (Behjati & Tarpey, 2013; Liu et al., 2012; Reis-Filho, 2009)

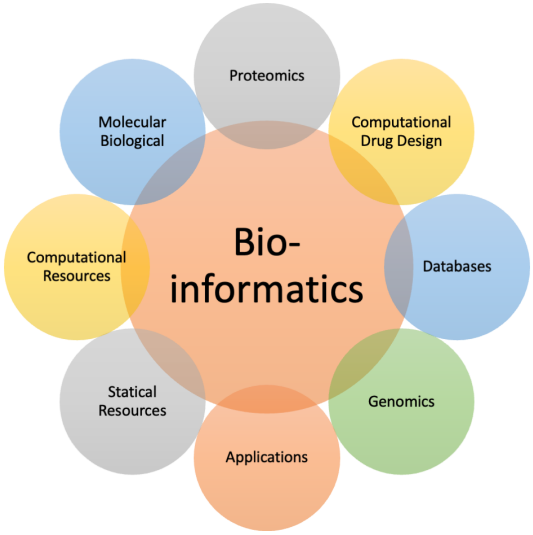

Figure 1.2: The meaning of Bio-informatics research in related fields

1.2 Neurodegenerative diseases

Neurodegenerative diseases are threatening human life expectancy. As the condition worsens, patients develop complications such as swallowing disorders, bladder and bowel dysfunction, blood pressure fluctuations, and problems with sleep, breathing, cardiac function, memory, and cognitive abilities. The current treatment for these diseases is based on suppressive therapy, mainly because there are no specific drugs to cure these diseases. (R. C. Brown et al., 2005; Checkoway et al., 2011; Forman et al., 2004; Hung et al., 2010; Terreros-Roncal et al., 2021) The pathology remains unclear among well-known diseases, such as Alzheimer's, Parkinson's, and Prion diseases; These diseases have in common that biological experiments often reveal distinctive responses that can be used to determine whether a lesion has occurred. However, the process is arduous to observe using traditional biological methods because the transformation of such

critical structures often takes place in milliseconds. Abnormal secondary structural changes are often a recognizable feature of these diseases and are usually accompanied by more pronounced signs of aggregation. In this case, computer simulations can simulate conformational changes in processes that cannot be directly observed. More evidence about neurodegenerative diseases can be obtained by representing the structural changes of such features, which is crucial for studying pathology and drug development. (Mandaci et al., 2020; Murphy, 2002)

1.2.1 Alzheimer's Disease

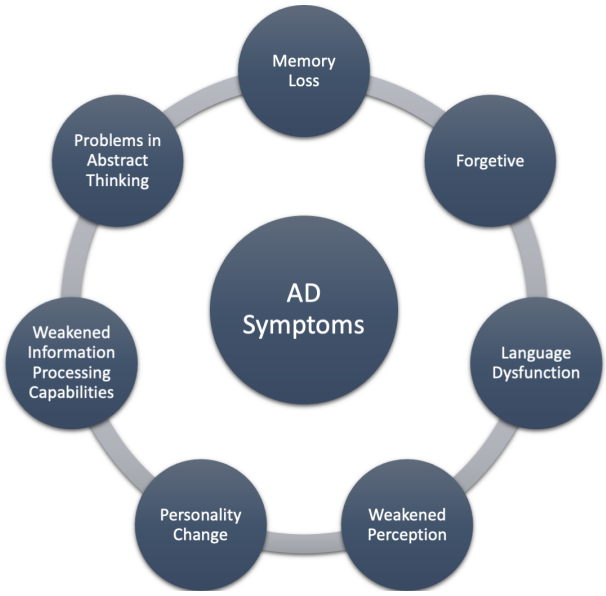

Figure 1.3: The major symptoms of Alzheimer's Disease patients.

Alzheimer's disease(AD) is currently proposed to pose one of the greatest threats to humans in the coming decades and is characterized by unclear pathogenesis. AD patients suffer from neural disorders shown in Fig. 1.3. Uncertainty about the cause of the disease makes it incurable. Typically, the main features that define Alzheimer's disease are abnormal secondary structural changes and amyloid-β aggregation. This lesion begins with conformational change, and

then a chain reaction leads to lesions in a series of normal amyloid-β molecules. However, the conformational change of the first amyloid protein is the most critical question and remains unclear. (Barz et al., 2018; Finder & Glockshuber, 2007; Hane & Leonenko, 2014; Luo et al., 2002; Yoshiike et al., 2001) A growing number of research groups have proposed hypotheses for AD drug candidates. These drugs are often studied only for some aspects of their pathological characterization, and it is difficult to describe their inhibition process clearly. If we want to test the inhibition directly of these drugs, they may cost a lot of effort and money and may fail. Research of these diseases is imminent, but without efficient and practical methods to model conformational changes in the secondary structure of target proteins, modeling the pathological process of Alzheimer's disease will be challenging. The conformational change process remains challenging for achieving a pathological understanding of neurodegenerative diseases and simulating potential responses to drug candidates.

1.2.2 Prion Disease

Prion disease is a progressive, fatal, incurable degenerative brain disease. In humans, prions are believed to be the cause of Creutzfeldt–Jakob disease (CJD), its variant (vCJD), Gerstmann–Sträussler–Scheinker syndrome (GSS), fatal familial insomnia (FFI), and kuru; in animals, Prion diseases are also prevalent, it was first identified in sheep and later in various mammals. The historically horrific Bovine Spongiform Encephalopathy (BSE), also named the mad cow disease, is typical. (n.d.)

Prion disease is a protein disease or a disease of structurally abnormal proteins. Similar to Alzheimer's disease, Prion disease is associated with the misfolding of specific proteins, the protein that prions are made of (PrP). Once a PrP starts to misfold, it is often irreversible, and its misfolding is considered to be infectious. The standard form of the protein is called PrP C (cellular), while the infectious

form is PrP Sc (scrapie). when a beta-structure-rich insoluble conformer (PrP Sc) approaches an alpha-helix rich prion protein (PrP C), the latter is "infected" by the former and starts to undergo similar structural changes to achieve the transformation from PrP C to PrP Sc. (Féraudet et al., 2005; Khalili-Shirazi et al., 2005; Tuo et al., 2001) Eventually, many PrP Sc structures will aggregate. Prion aggregates are stable, and this structural stability means that prions are resistant to denaturation by chemical and physical reagents, which means they cannot be destroyed by ordinary sterilization or cooking. Proteases cannot degrade them. Coincidentally, prion diseases and other neurodegenerative diseases face similar dilemmas as Alzheimer's. Similarly, there is no complete cure for prion diseases. The role of PrP in the organism is not clear. (Abbott, 2010)

1.3 Dissertation outline

The overall theme of this dissertation is to develop novel strategies to validate neurodegenerative diseases by utilizing computational tools, to understand the proteins' dynamics transformation, hence revealing the cause of the conditions. In subsequent sections, we focus on two representative NDs, AD and Human Prion Disease, for which we used computer simulations on them and analyzed their pathology based on the simulation results. We will describe in detail the computer simulation resources we used and the specific steps we took.

Chapter 2

Computational Approaches for Alzheimer's disease

In this chapter, computational approaches including protein-ligand docking, molecular dynamic simulations and model generations are applied to Alzheimer's disease(AD) and its protein causative agent, amyloid beta(Aβ). we carried out several comprehensive dynamic simulations to investigate the structural change of Aβ. The misfolding of Aβ by multiple causes were dynamically revealed. Compared to previous studies, the innovations of this thesis are:

- The results of molecular dynamics simulations dynamically demonstrate that misfolding occurs in Aβ and more than one pathogenic pathway was revealed.
- Related Herpes Simplex Virus type 1(HSV-1) to AD and Aβ . The glycoprotein B of HSV-1 was identified to be a misfolding inducer candidate.

2.1 Introduction

2.1.1 Background of Alzheimer's Disease

Alzheimer's disease is a progressive neurologic disorder that causes the brain to shrink (atrophy) and brain cells to die. Alzheimer's disease is the most common cause of dementia — a continuous decline in thinking, behavioral and social skills that affect a person's ability to function independently. Approximately 5.8 million people in the United States age 65 and older live with Alzheimer's disease. Of those, 80% are 75 years old and older. Out of the approximately 50 million people worldwide with dementia, between 60% and 70% are estimated to have Alzheimer's disease. The early signs of the disease include forgetting recent events or conversations. As the disease progresses, a person with Alzheimer's disease will develop severe memory impairment and lose the ability to carry out everyday tasks.

Medications may temporarily improve or slow the progression of symptoms. These treatments can sometimes help people with Alzheimer's disease maximize function and maintain independence for a time. Different programs and services can help support people with Alzheimer's disease and their caregivers. There is no treatment that cures Alzheimer's disease or alters the disease process in the brain. In the advanced stages of the disease, complications from severe loss of brain function — such as dehydration, malnutrition, or infection — result in death.

2.1.2 Alzheimer's Disease in History

Alzheimer's disease is named after a doctor - Alois Alzheimer, a young psychiatrist in his late 30s, was a hard-working clinician dedicated to understanding the relationship between brain disorders and mental illness. In 1901, he received a patient in her 50s named Auguste Deter, who was initially found to have vari-

ous memory problems, and then she became progressively more unstable and was admitted to a psychiatric hospital until her death. After Dr. Alzheimer microscopically examined Ms. Deter's brain with new stains, he found that they revealed the presence of what we now call amyloid plaques and neurogenic fiber tangles. Strange as it may seem now, the original report of Alzheimer's disease in 1906 related this particular brain pathology to a clinical syndrome. (n.d.) At that time, patients were found to have similar symptoms of dementia, but they were not well differentiated. Subsequently, the number of cases of Alzheimer's disease began to increase dramatically and showed a continuous growth trend.

2.1.3 Explosive growth trend

Alzheimer's disease is the sixth leading cause of death in the United States population and the fifth leading cause of death in adults older than 65. The primary risk factor for AD is age, but it also depends on race and ethnicity. In 2020, an estimated 5.8 million Americans aged 65 years or older had Alzheimer's disease. This number is projected to nearly triple to 14 million people by 2060. In 2010, the costs of treating Alzheimer's disease were projected to fall between $159 and $215 billion. By 2040, these costs are projected to be between $379 and more than $500 billion annually. Death rates for Alzheimer's disease are increasing, unlike heart disease and cancer death rates, which are on the decline. Dementia, including Alzheimer's disease, is under-reported in death certificates; therefore, the proportion of older people who die from Alzheimer's may be considerably higher. (Broe et al., 1990; De Gage et al., 2014; Hurd et al., 2013; Matthews et al., 2018; Mohelska et al., 2015; Nedelec et al., 2022; Sosa-Ortiz et al., 2012)

2.1.4 Main features for Alzheimer's disease

Alzheimer's disease has been known for decades, but no cause or effective treatment has been found so far (Burns & Iliffe, 2009), which is inseparable from

the characteristics of Alzheimer's disease itself. Alzheimer's disease is a chronic and progressive disease whose primary symptom is dementia and will gradually worsen over several years. Because the disease mainly occurs in people over 65 years of age, and the early symptoms of Alzheimer's disease are mild, it is not easily detected, and once diagnosed, there is no effective treatment, only medication to slow down the aggravation of the disease. (Scheltens et al., 2016; Selkoe, 1991)

2.1.5 Alzheimer's disease and Amyloid-β

After years of attempted studies, one of the most commonly accepted causes of Alzheimer's disease is now believed to be Amyloid-β (Aβ). Neuropathological deposition of Aβ in senile plaques and brain amyloid angiopathy associated with dementia in FAD and sporadic AD. (Hunter et al., 2016)

2.2 Aβ simulations

Amyloid beta (Aβ) is a peptide of 36-43 amino acids and is the main component of the amyloid plaques found in the brains of Alzheimer's patients. Peptides are derived from amyloid precursor protein (APP), cleaved by β-secretase and γ-secretase to generate Aβ in a cholesterol-dependent process presented as a substrate. Aβ molecules can aggregate to form flexible soluble oligomers that may exist in several forms. It is now believed that specific misfolded oligomers (called "seeds") can induce other Aβ molecules to take the form of misfolded oligomers, leading to a chain reaction similar to that of prion infection. This oligomer is toxic to nerve cells. Another protein implicated in Alzheimer's disease, tau, also forms this prion-like misfolded oligomer, and there is some evidence that misfolded Aβ can induce tau misfolding. Previous scientific research identified that Aβ40 or Aβ42 are the best candidates that are regarded to be

highly related to AD formation; Specific ratio Aβ4042/Aβ4040 can predict amyloid-PET status. (Doecke et al., 2020)

2.2.1 Aβ Monomer and Dimer

Some misfolded oligomers of Aβ (called "seeds") can also induce other Aβ molecules to take the misfolded oligomeric form, and these oligomers are toxic to neuronal cells. This mode of transmission is very close to that of prions. Therefore, studies targeting Aβ monomers are necessary. Monomers Aβ can be roughly divided into two categories: those that do not possess a misfolded structure; and those that have produced a misfold. Here we have picked an unowned error fold Aβ40 as a singleton (PDB ID: 2LFM), as shown in 2.2. These

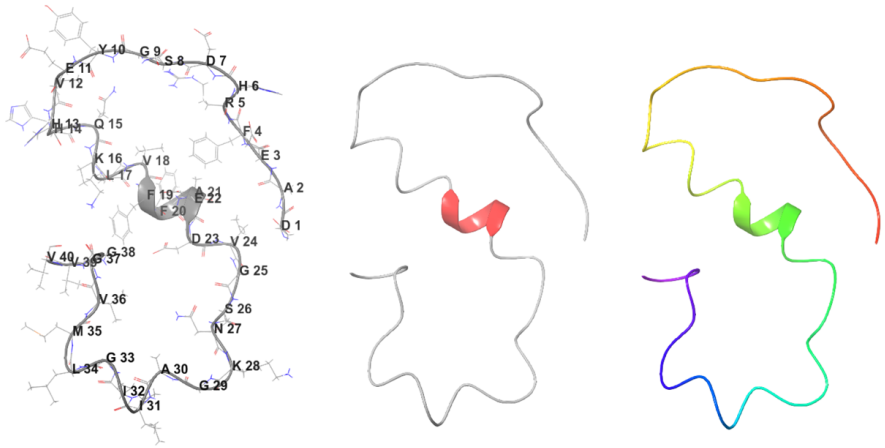

Figure 2.1: Three operation methods for Aβ40 structure(PDB:2FLM)

figures show three different presentations; the left one shows the position and number of each amino acid; the middle one shows the secondary structure of this protein (red for alpha helix, gray for loop); the right image shows the rainbow diagram of this protein from the N-terminal to the C-terminal (red for the N-terminal position, purple for the C-terminal).

The flexibility of Aβ is high because there is no sufficiently abundant secondary

structure or clear internal conditions of action of the protein sequence. After constructing a water solubility model for it, the basic spatial morphology changes significantly in a simulated environment of fewer than five ns. This is consistent with the original nature of Aβ, which has good water solubility in its standard form and is usually circulated in the blood, cerebrospinal fluid, and interstitial fluid in the human body. Therefore, there is no apparent progress and significance of computer simulation studies for Aβ.

Unlike the monomer model, the dimer model has its own advantages in various

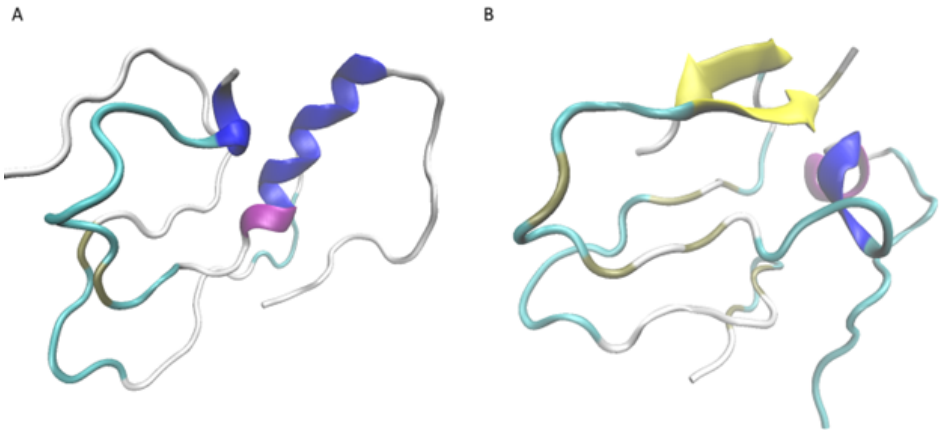

Figure 2.2: Simulation of the conversion of the secondary structure of amyloid-beta. Structures A and B are the structures of two amyloid beta before and after MD simulation. The initial complex structure was generated by protein-protein docking. The yellow parts represent beta-sheets; the Blue and purple parts represent the alpha-helix and pi-helix

aspects: firstly, due to diploid binding, the two proteins interact and bind to each other, maintaining the spatial structure better than the haploid; secondly, the diploid can be shaped in various combinations, i.e., a combination of one wild type and one pathotype, or a variety of two wild types; finally, we can analyze the structure of the diploid model on this basis, and then further combine other binders. We applied protein-protein docking to dock misfolded amyloid-beta on different surfaces of normal amyloid and tested different environmental conditions to examine their effects on converting secondary structures.

The preliminary study only presents one of the possible models for 2 Aβs.

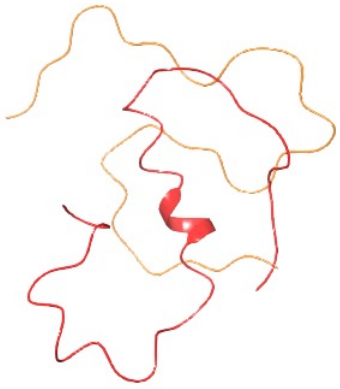

Figure 2.3: Dimer model of Aβ in pure cartoon style. The red chain represents the unmisfolded Aβ, and the orange chain is the misfolded Aβ

Protein-protein docking also shows the ranking of different relative positions. Then the created models were preprocessed and transferred to GROMACS to run the MD simulations. According to our preliminary study, the simulation will be operated in normal conditions, which will be required to trigger the secondary structural change and the increasing beta sheets. GROMOS96 54a7 force field will be used as the force field for operating MD simulations and other projects in this proposal without a specific state for using other force fields. By comparing these models' performance, we will extend the explanation for proteopathies and show the relationship between the 2 Aβs more comprehensively.

2.2.2 Docking compounds identification

In our previous work, drug screening was applied to find possible drugs that perform better in inhibiting Aβ aggregation. We built pharmacophore models

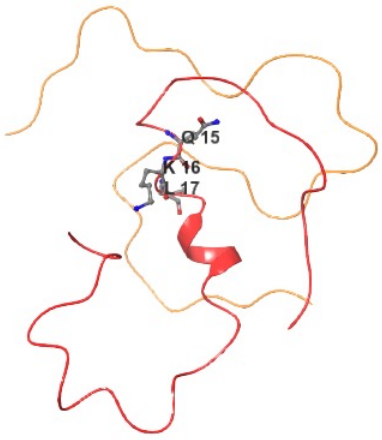

Figure 2.4: Dimer model of Aβ in pure cartoon style with highlighted 3 amino acids. The red chain represents the unmisfolded Aβ, and the orange chain is the misfolded Aβ

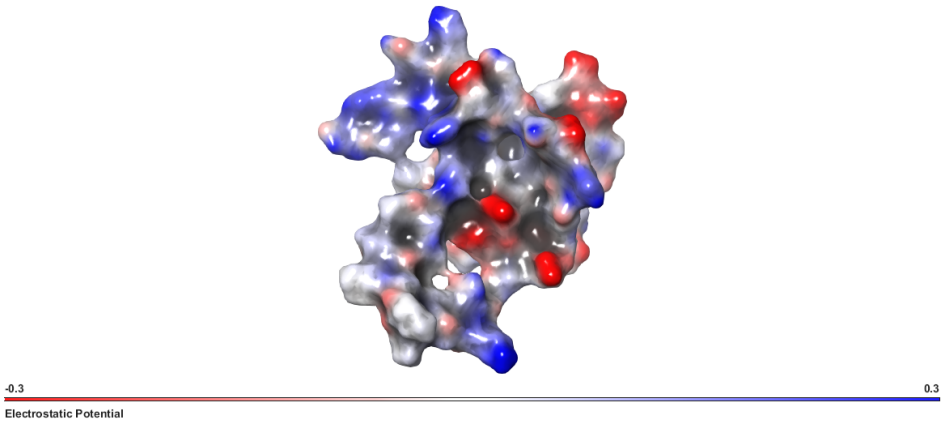

Figure 2.5: Dimer model of Aβ showing protein surface. The domain was colored in red and blue to reflect the local charge of that region; red means a negative charge, and blue means a positive charge

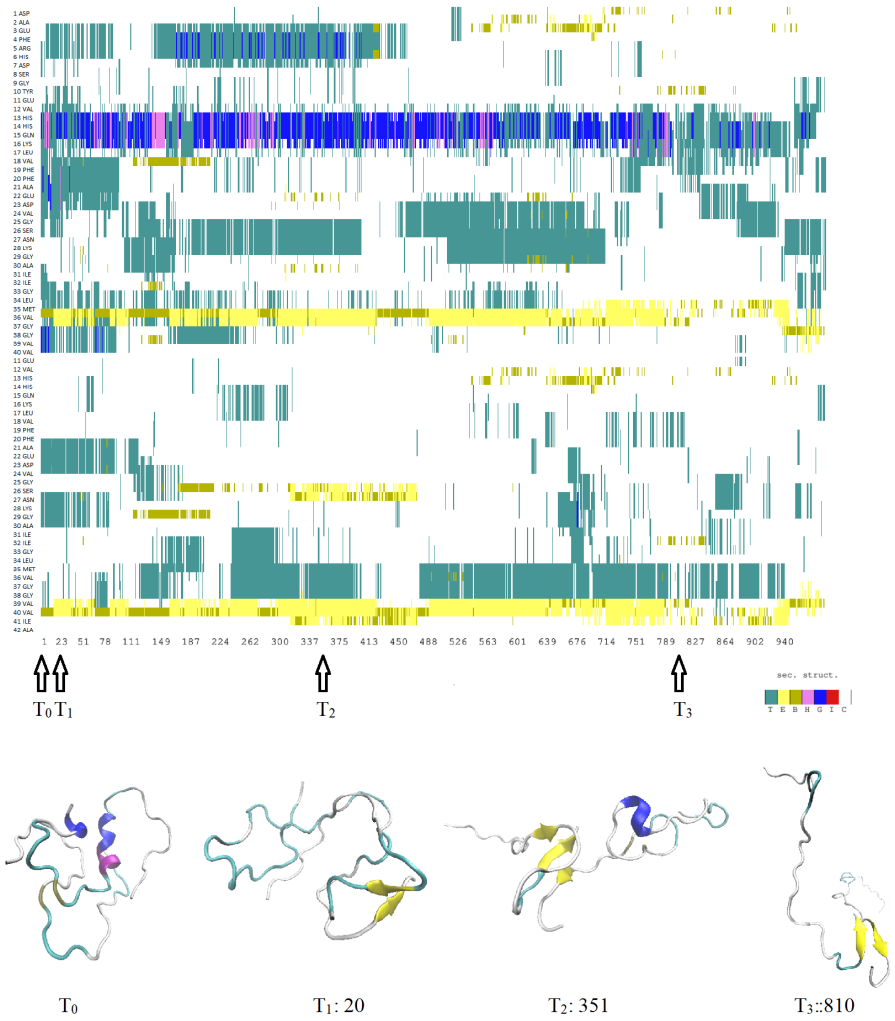

Figure 2.6: MD simulation trajectory secondary change for the dimer model. he upper figure reveals the secondary change for specific Amino acids(Y-axis) with simulation frames(X-axis); The yellow pattern means beta-sheets. each frame is recorded every 0.1 ns. The figures below reveal the conformation of the complex in certain frames.

and then performed screening work against online drug databases. Drug repurposing is also an excellent way to find inhibitors because the toxicity of the drugs has been well studied. The approved (1,516 drugs) and experimental (4,788 drugs) drug databases were used primarily for drug screening as they will provide less toxic drug candidates. We will then expand the database to include ligands from the PDB database (19,500 drugs), the ChEBI database (27,950 drugs), and the HMDB database (39,060 drugs) to obtain a wide range of compounds to be tested. After finding the binding compounds, we performed a binding site analysis to highlight the amino acids that constitute the binding pocket, build a pharmacodynamic model, and finally used the model to search through the zinc drug database (10,639,400 compounds). The screening database can be found in SwissSimilarity. At the end of the screening, the top-ranked candidate compounds were collected and analyzed for binding affinity. The offline chemical synthesis was applied by manual modification using standard features from the top candidate compounds.

Miquelianin (quercetin 3-O-glucuronide), shown as figure 2.8, was screened

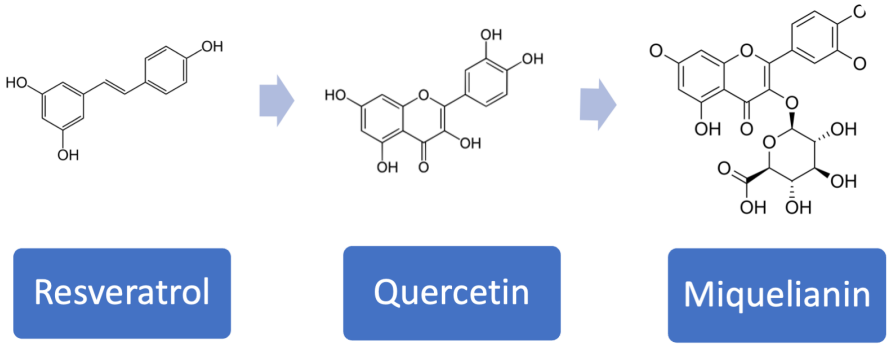

Figure 2.7: The screening procedure that leads to the identification of Miquelianin. The virtual screening starts with Resveratrol; Quercetin was screened out as an intermediate candidate; Miquelianin was found from Quercetin.

out as the best candidate on the binding site of Aβ dimer.

Miquelianin is a phenolic compound it is found in wine. Its medicinal value

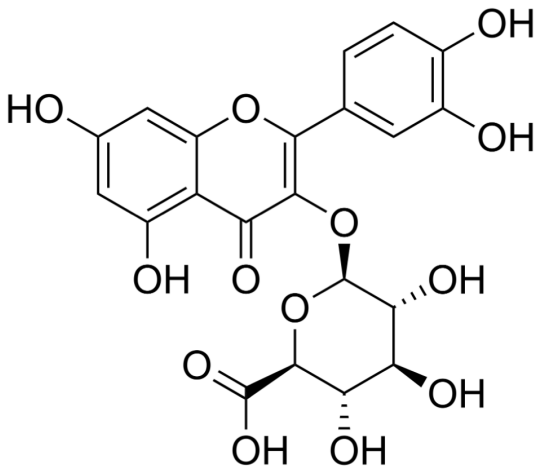

Figure 2.8: 2-D structure of Miquelianin

is not yet clear, but past experiments have shown that it can be absorbed by the small intestine and reach the central nervous system. (Ghiselli et al., 1998; Juergenliemk et al., 2003) With the help of protein-ligand docking, we acquired the binding model of the Aβ dimer with Miquelianin on it, and it is shown as Fig2.9. The Aβ dimer was set as a receptor mentioned before. Miquelianin can bind with Aβ dimer on the unmisfolded side, close to the pH-sensitive 3 amino acids.

To understand what significance Miquelianin binding has, we performed MD simulations for its model. The simulation results are shown in Fig. 2.10. The

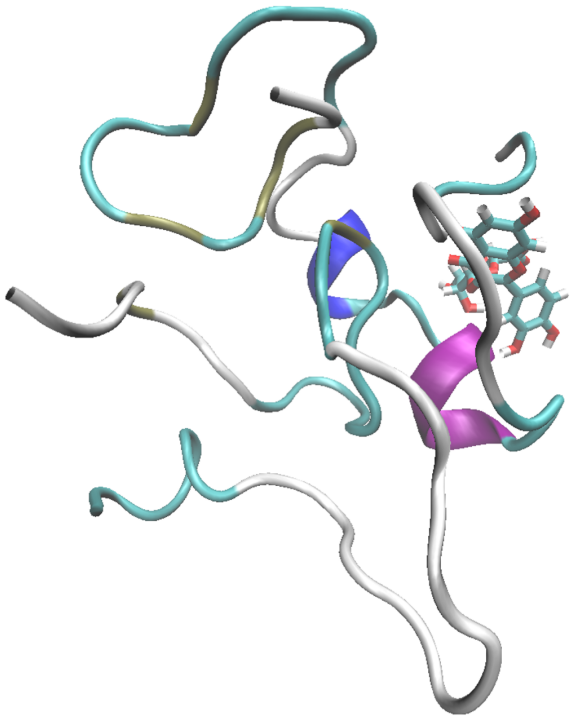

Figure 2.9: Aβ dimer model binds with Miquelianin. Aβ is shown by cartoon style, and Miquelianin is shown as a balls and bonds structure. We can see the

binding of Miquelianin significantly delays the formation of beta sheets in Aβ Dimer without changing any simulation conditions. Even if beta sheets are formed, their stability is reduced.

2.2.3 Heparin

Heparin is a drug and a naturally occurring glycosaminoglycan. Because heparins depend on antithrombin activity, they are considered to be anticoagulants. Natural heparin is a polymer with a molecular weight range of 3 to 30 kDa, although most commercial heparin preparations have an average molecular weight of 12 to 15 kDa. Heparin is a member of the glycosaminoglycan family of carbohydrates (including the closely related molecule acetyl heparan sulfate) and consists of variable sulfated repeating disaccharide units. The central disaccharide units that occur in heparin are shown below. (The most common disaccharide unit) consists of 2-O-sulfated aldohexanoic acid and 6-O-sulfated, N-sulfated glucosamine, IdoA(2S)-GlcNS(6S).

We created models for Aβ dimer binds with heparin. This process was mainly done manually due to the limited resources to achieve glycosome docking. We tried to use protein-ligand docking regarding the glycosome as regular ligands, but we failed because the algorithm can not include the consideration of glycol groups. The output broke the reasonable 3-D structure of heparin. Thus, we manually moved the heparin to the Aβ dimer binding pocket and changed the binding orientation to ensure they were in the proper range. After that, we conducted several rounds of MD simulation for them, and only one group showed positive results: the group of heparin terminated with IdoA(2S)-GlcNS(6S) close to Aβ was not getting apart from the binding site after MD simulation. All the other groups showed typical results: heparin gets out of the binding pocket immediately after the start of MD.

Interestingly, for the good binding group, after introducing heparin into the system, the parallel structure of Aβ is not disrupted with the onset of MD when

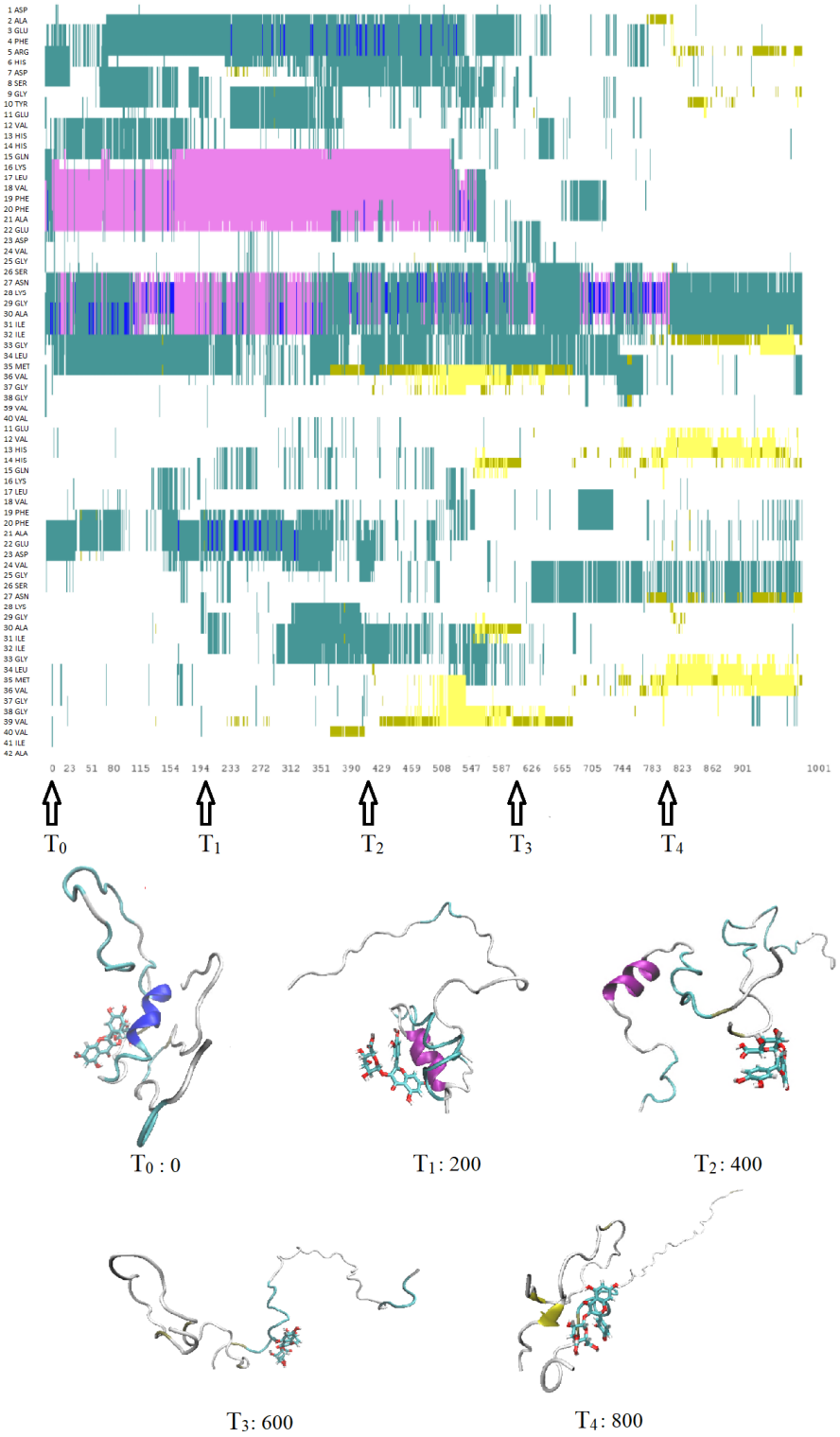

Figure 2.10: MD simulation trajectory secondary change for the dimer model binds with Miquelianin. The upper figure reveals the secondary change for specific Amino acids(Y-axis) with simulation frames(X-axis); The yellow pattern means beta-sheets. each frame is recorded every 0.1 ns. The figures below reveal the conformation of the complex in certain frames.

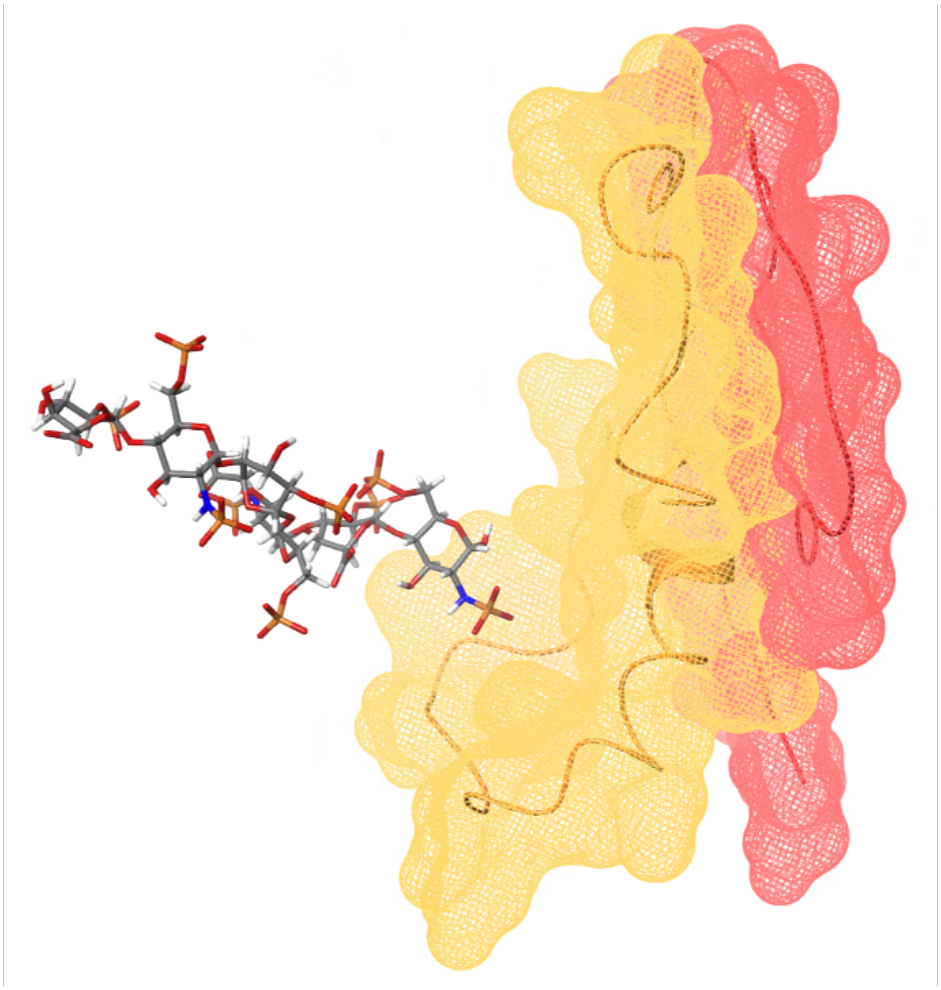

Figure 2.11: Binding model of Aβ and heparin. The Aβ dimer is shown in pure cartoon style with a mesh surface highlighted in different colors for 2 entries. The red chain represents the unmisfolded Aβ, and the yellow chain is the misfolded Aβ. IdoA(2S)-GlcNS(6S) binds to the dimer

heparin is terminated with IdoA(2S)-GlcNS(6S) close to Aβ, which provides the possibility of fibrillation of Aβ. current misfolding infection theory proposed the one-by-one misfolding infection, which means the aggregation is done layer-by-layer. Our simulation results reveal the process of this misfolding transfer, but it requires the parallel spatial location of each Aβ. Hence, heparin's function here is pretty solid for creating an appropriate binding model for further misfolding aggregation.

2.3 Herpes Simplex Virus type 1(HSV-1) and Alzheimer's disease

Since 2017, the relationship between HSV-1 and AD has been emphasized, there are more and more papers related to it, and a large amount of evidence of HSV1 action has been linked to AD, and we also hope to find a direct relationship between the two using protein structure simulation.

2.3.1 Herpes Simplex Virus Type 1

Herpes simplex virus(HSV)- commonly known as herpes- can be divided into two types: herpes type 1 (HSV-1, or oral herpes) and herpes type 2 (HSV-2, or genital herpes). Most commonly, herpes type 1 causes sores (sometimes called fever blisters or cold sores) around the mouth and lips. HSV-1 can cause genital herpes, but most cases of genital herpes are caused by herpes type 2. Studies have shown that HSV-1 or HSV-2 infection can lead to meningitis (inflammation of the brain and spinal cord coverings) or encephalitis (inflammation of the brain). (Tyler et al., 2004)

Besides the protein-protein direct infection, preliminary studies show the HSV-1 will also be able to trigger Aβ to get misfolded; Thus, validation for the HSV's effect on making wildtype Aβ get misfolded is essential to construct the path-

way of the misfolding process of the very first Aβ. We followed the homologous amino acids' strain and conducted protein-protein docking for more relative amino acids' fragments to glycoprotein B.

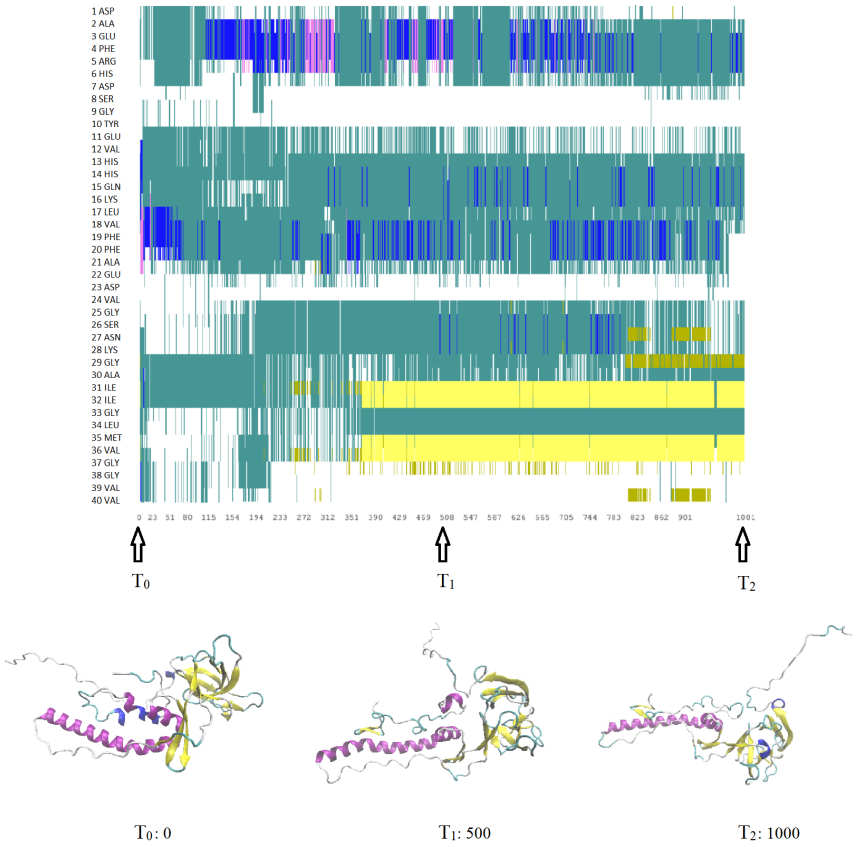

Figure 2.12: Dimer model of Aβ binds with HSV-1 Glycoprotein B. The upper figure reveals the secondary change for specific Amino acids(Y-axis) with simulation frames(X-axis); The yellow pattern means beta-sheets. each frame is recorded every 0.1 ns. The figures below reveal the conformation of the complex in certain frames.

2.4 Methodology and model generation

In this chapter, the PDB files used are all from the RCSB Protein Data Bank ($https://www.rcsb.org/$). The first frame is selected by default when the PDB contains multiple frames to build the model. Maestro by Schrödinger is

used to process and optimize PDB files, which mainly includes: the deletion of unwanted interfering factors in the file, such as redundant small molecules or water molecules; pre-processing of proteins; spatial and energetic optimization of branched chains; and repair of incorrectly written protein sequences. To estimate the interactions between target proteins and small molecules, we conducted ligand-protein docking by using the Ligand Docking panel in Maestro. Before running docking jobs, a receptor grid box was generated based on the existing ligand in the protein structure. The size of the receptor grid box was set as default (20 rÅ). Ligand–protein docking was performed in extra-precision (XP) mode. To further investigate the dynamic interactions between protein and small molecules, we conducted molecular dynamics (MD) simulations by using GROMACS version 2018.1 and CHARMM36 force field. The starting coordinates of the protein-ligand complex were obtained from a ligand-protein docking study. Then, we used CHARMM-GUI to build the MD simulation solution box, a cubic box with a length of 125 , which was then filled with water. Next, the minimized structures were equilibrated using an NVT ensemble (constant Number of particles, Volume, and Temperature) and an NPT ensemble (the Number of particles, Pressure, and Temperature). The target equilibration temperature was 300 K. Finally, MD simulations were performed for 100 ns. After the MD simulations, we calculated the root-mean-square deviation (RMSD) and the potential energies.

The Detailed scripts and program commands are listed in Chapter 4.

Chapter 3

Computational Approaches for Human Prion Disease

In this chapter, protein structure prediction, molecular dynamic simulations and protein-protein docking were conducted to the Prion Protein (PrP) to observe the dynamic structural change of Prion Protein (PrP). We elucidated that the N-terminus of PrP is the control of the whole PrP misfolding process. Compared to previous studies, the innovations of this chapter lines in:

- PrP whole sequence models are generated with different orientations, including the missing N-terminus and C-terminus, which makes the MD simulation results different from previous simulations that only used resolved domain.
- MD simulations reveals the misfolding trends for models. One of these models is particularly prone to its own misfolding.
- Based on the simulation results, a novel hypothesis for the initiation mode of PrP misfolding is proposed.

3.1 Introduction

3.1.1 Background of Prion Diseases

Prion diseases usually refer to a group of progressive neurodegenerative diseases known as transmissible spongiform encephalopathies (TSEs)(Prusiner, 1991; Prusiner & Hsiao, 1994; Prusiner et al., 1998). Similar to Alzheimer's disease, the characteristics of the disease include a long latency period, a greatly affected nervous system, and irreversibility, making these disorders very dangerous.(Linden et al., 2008)

The pathogen of prion diseases is prion protein(PrP), which is an endogenous protein. Similar to Aβ, prion protein becomes infectious after misfolding for unknown reasons and becomes a particular protein capable of inducing misfolding of the normal prion protein. The function of prion proteins is currently unknown, and studies have shown that they are usually widely distributed in the brain and are heavily aggregated in the brains of patients with prion disease.(Silveira et al., 2005; Westergard et al., 2007; Wulf et al., 2017)

3.1.2 The History of Prion Diseases

In the 1960s, British biologist Alps used radiation treatment to destroy DNA and RNA, but the tissues were still infectious. He thought the causative agent of "sheep itch" was not nucleic acid but probably protein. In 1947, it was discovered that mink had the cerebral softening disease, and its symptoms were similar to those of "sheep scratching." Later, chronic wasting disease (atrophy) in horses and deer and spongiform encephalopathy in cats were discovered. The most shocking was the unprecedented panic caused by the "mad cow disease" in the UK and the world in the spring of 1996, which triggered political and economic turmoil and made people "fearful of talking about cattle." In 1997, the

Nobel Prize in Physiology and Medicine was awarded to Stanley B.P Prusiner, an American biochemist, for his discovery in 1982 of a new type of organism - A prion. Before that, it used to have many names, such as unusual viruses, lentivirus, and infectious brain-like changes. (P. Brown et al., 1998; Chesebro, 2003)Over the years, many experimental studies have shown that a group of protein particles with a molecular weight of 27,000 to 30,000 cannot be traced to any nucleic acid are highly resistant to various physical and chemical effects and are highly infectious. It is a specific cause of the infectious encephalopathy TSE in humans and animals. (Daskalov & Saupe, 2015; Polo, 2000; Poser, 2002) Prions have gone beyond the biological concept of classical virology. Research has shown that the idea that proteins undergo mutations or conformational changes under specific conditions and change from benign to malignant, i.e., become infectious protein particles, presents a solid challenge for the conventional view. Diseases caused by prions have been discovered and confirmed. The incidence of both the population and animal populations is on the rise worldwide, so the study of prions has excellent theoretical and urgent practical significance. Its theoretical significance lies in the opening up of a new field of biomedical research, which requires the close cooperation of multiple disciplines such as cell biology, molecular biology, molecular genetics, protein chemistry, molecular virology, neuropathology, etc., to answer a series of questions brought about by prions; its practical significance lies in the development of accurate and reliable diagnostic techniques, comprehensive monitoring and detection of prion diseases, especially to achieve the early prevention of mad cow disease and medically-derived early prevention of infection. For humans, Creutzfeldt-Jakob disease (CJD) is a disease caused by prion proteins. As with all other transmissible spongiform encephalopathies, it is currently incurable. (Dextera & Jenner, 2013; Kovacs & Budka, 2008; Wadsworth & Collinge, 2010)

3.2 Human Prion Protein

Prions are very common among animals, while for humans, although not so common, we should be especially cautious because of their incurable nature. In this chapter, human prion diseases are mainly referred to as Creutzfeldt-Jakob disease (CJD).

3.2.1 Human Prion Protein sequence

The protein we used for this chapter is the Human Prion protein(Uniprot ID: P04156 · PRIO_HUMAN)(Consortium, 2015), NMR resolved structure was acquired from RCSB (PDB ID: 2LSB)(Biljan et al., 2012). Like the Blast result shows in Fig.3.1(Ye et al., 2006), prion proteins also have a high degree of homogeneity across species. This provides a theoretical basis for the cross-species transmission of prion diseases. Bovine spongiform encephalopathy (BSE), also known as "mad cow disease," was a serious public health safety incident in the United Kingdom in the 1980s. The main way to infect people is through human consumption of meat products from cattle with BSE. (Ireland, 2003; Wilesmith et al., 1988)

Human PrP is a protein consisting of 253 amino acids, as shown in Fig.3.2. Clinical studies have shown that it can induce the prion protein, which is a typical structure on the nerve cell, to change into a wrong structure and undergo an aggregation reaction, and through this mechanism, a new prion protein is introduced, which continuously replicates itself and passes to neighboring cells, and eventually spreads throughout the brain. (Ma & Lindquist, 2002)

3.2.2 Structure Prediction for N-terminus

Scientific evidence suggests that the N-terminus of PrP plays a significant role in determining its secondary structure and misfolding process. We must con-

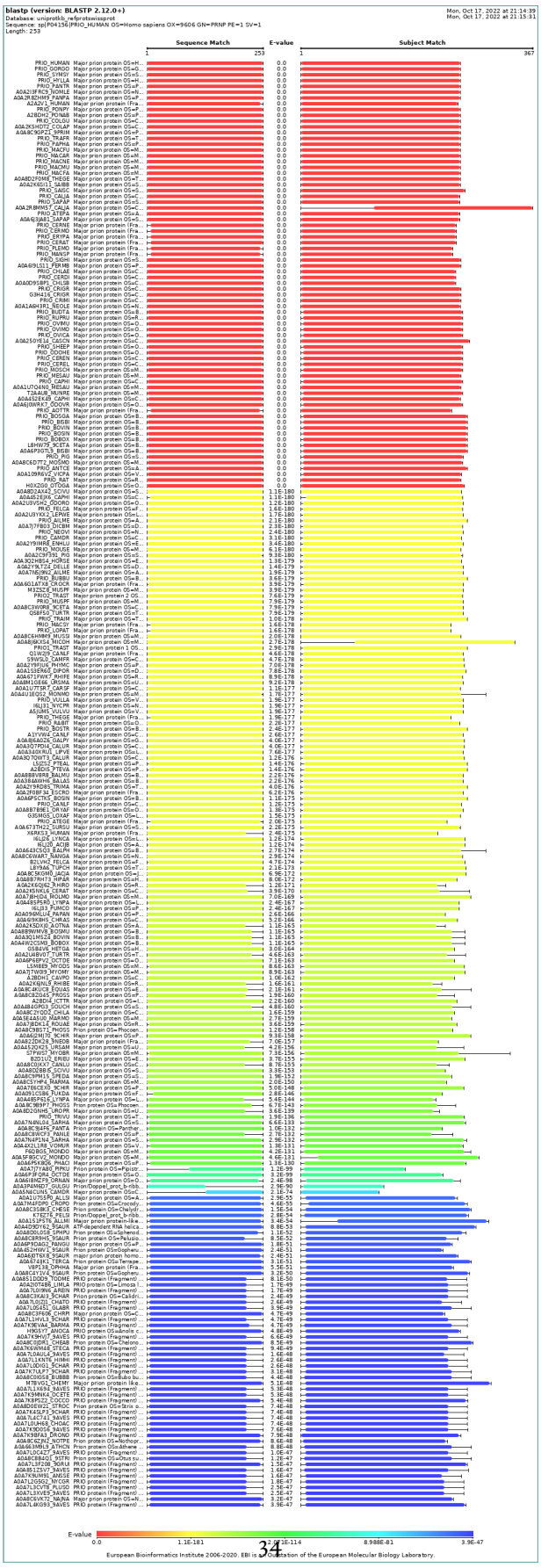

Figure 3.1: The Blast Results for Human Prion protein. The results are ranked in the order of identity.)

Figure 3.2: The prion protein sequence with highlights. In the upper part of the figure, polar amino acids are underlined in red; in the lower part, hydrophobic amino acids are underlined in blue.

struct the full-length protein structure to understand how N-terminus affects the whole protein structure from the protein structure level. With the help of the RoseTTA algorithm, we initially achieved this. We placed the sequences of PrP into the Robetta server built, and the predicted model is shown in Fig.3.3. (Das & Baker, 2008; Kaufmann et al., 2010; Park et al., 2018) We can clearly see that the two models have a highly overlapping domain because the RosaTTA algorithm scans the placed protein sequences and then matches them within the protein database, and when there is a highly overlapping known structure, this part of the structure is referenced. For PrP, most of the available structures focus on the sequence of amino acids 120-225, so there is not much difference in this part of the structure; the main differences in structure prediction are in the nitrogen and carbon ends of the protein (the carbon ends are relatively much shorter and have a minimal decisive effect on the structure difference), which is in line with our experimental purpose. We can clearly see that under the predicted structure dominated by a large number of loops, the nitrogen terminus

of the protein presents two spatial positions relative to the known structure: on its left and upper side, respectively. To have a deeper understanding of the practical significance of these two models, we performed MD simulations for each of them under the same conditions. The MD frames are listed in Appendix Fig.**??** The error estimation results of the above two models are shown in Fig.3.4

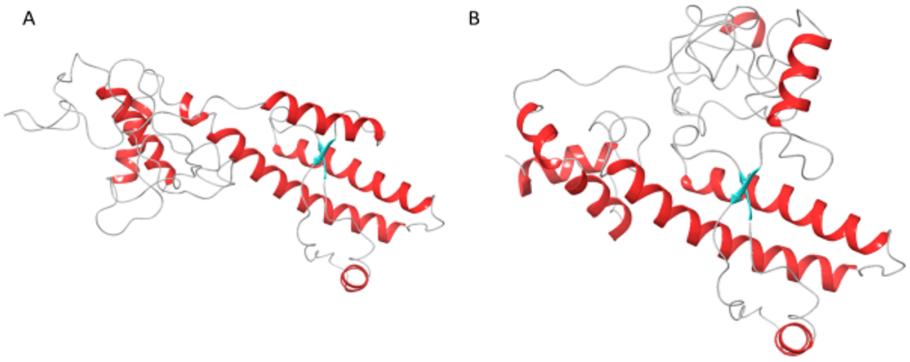

Figure 3.3: The 2 PrP models were predicted by using the RoseTTA algorithm. The proteins' backbone was presented using cartoon style with highlights of their secondary structures. Alpha helixes are colored in red, and beta sheets are colored in blue. Model A is the first candidate, while model B is the second.

The error estimates of model A and model B are shown in A and B, respectively. by analyzing this chart, we can find that the overall error estimate value of the existing domains is deficient, generally below 5, while the value of the nitrogen end is significantly higher than the carbon end, and all of them are higher than the known domains.

3.2.3 N-terminus Ions Chelate Hypothesis

Our preliminary study MD simulation results show supporting evidence that the N-terminus plays an essential role in PrP aggregation. Besides the direct impact from the N-terminus, recent research shows the full-length prion protein is easier to aggregate together than the globular domain only under the

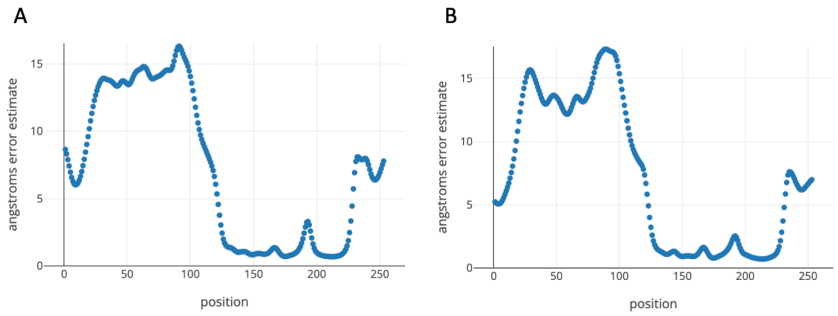

Figure 3.4: The error estimate charts for models in Fig3.3 The X-axis represents the number of the amino acid sequence, and the Y-axis represents the error estimate of the corresponding amino acid when predicting the model, the lower this value means, the more reliable the prediction result. We can see that the prediction accuracy of 120-225 amino acids is high, directly related to the fact that this domain has been acquired. The predicted results at the N-terminus are relatively the worst.

presence of some metal ions (Cu2+, Cu+, Zn2+). The hypothesis is proposing 4-Histidine-metal ions chelate. The model of the chelate is shown in Fig.3.6 (Schilling et al., 2020)

The main challenge is that copper ions are not adopted by default in several different MD simulation resources. This gives excellent difficulty and challenge to the simulation. Although they both have some packages provided by third-party developers, they both have conflicting versions in use. The solution we give is to use the included initial zinc ion as a model and change the ion radius(From 74 pm to 74 pm) and nomenclature to simulate the properties of the natural copper ion as much as possible since it has the same number of charges. However, the copper ions simulated according to the above-mentioned method could not play a role in stabilizing the chelate. In a concise time (less than 5 ns), the

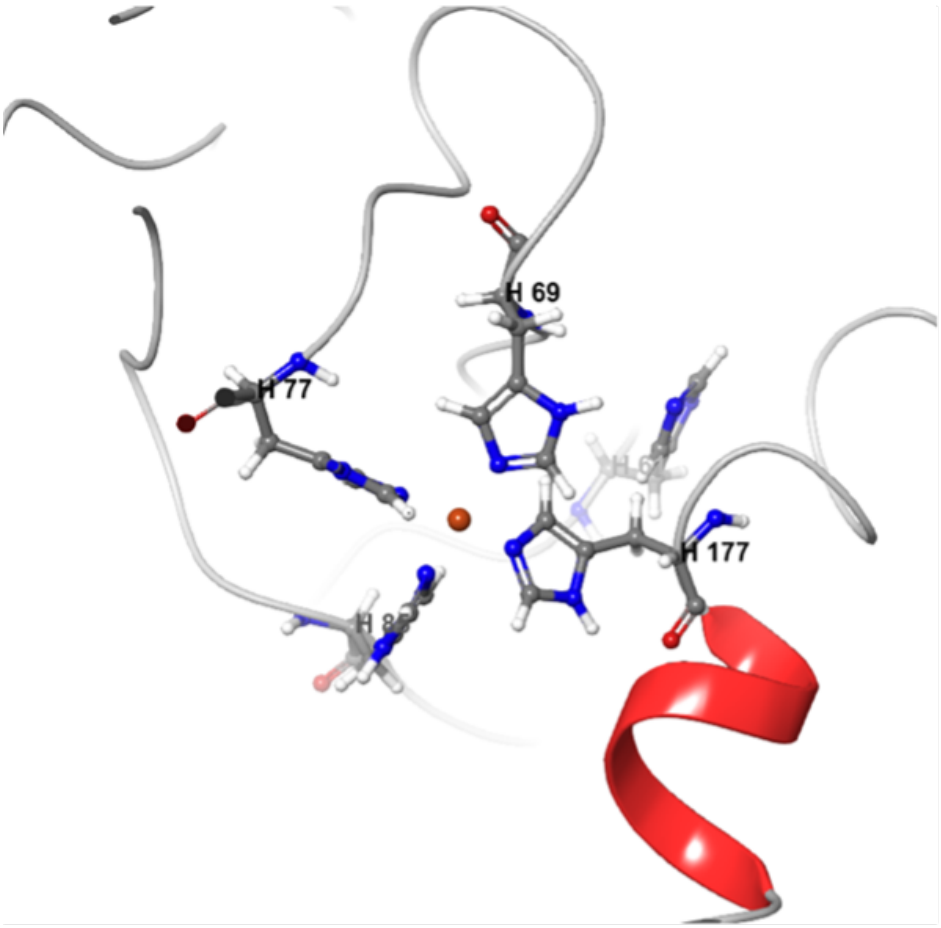

Figure 3.5: The model of 4-Histidine-metal ions chelate. This is a proposed model that formed with 4 Histidines and a Cu2+. The hypothesis proposed that this Cu2+ performs like a wedge: once it is removed, the chelate will be broken, and the full protein structure will change.

chelate structure spreads out, and the simulated copper ions leave their original positions in a disorderly motion similar to Brownian.

3.3 A Brand New Hypothesis on N-terminus

Admittedly, the evidence found so far does not seem to identify how the secondary structure of the N-terminus of PrP can be stabilized, yet numerous studies, including the simulation results used within this thesis, highlight the

influence of the N-terminus on the region where misfolding initiation occurs. With our predicted two models of PrP by Robetta, the available simulation data tell us they are both possible in two different forms. The error estimates indicate no significant advantage or disadvantage between the two. Interestingly, model B's beta sheets have a significantly increasing trend compared to model A. By comparing the structural differences between models A and B, we can easily find that there is no significant difference between the two secondary structures, and the results dominated by loop give a great possibility of spatial transformation of the entire domain; the most crucial difference lies in the spatial location relationship between N-terminus and primary domain.

Based on the above results, we propose the following hypothesis: when PrP C is in vivo, as an endogenous protein, its N-terminus is in a stable form due to some factors (e.g., copper ion chelate model), and in such a stable state, the amino acid structure of N-terminus cannot move freely, and thus its overall protein structure is relatively stable. The initiation of the lesion lies in the disruption of the steady state, where the structure of the N-terminus starts to become loose and undergoes various spatially free movements due to specific external factors (e.g., removal of copper ions in the copper ion chelate). Due to its lack of stable secondary structure, the loop-dominated N-terminus undergoes significant spatial position changes. This process of location migration may be random. When the relative position of the N-terminus and primary domain changes to the state shown in Model B, N-terminus promotes the extension of beta folding of primary domain and strengthens the stability of this folding, which marks the formation of PrP Sc and the beginning of misfolding.

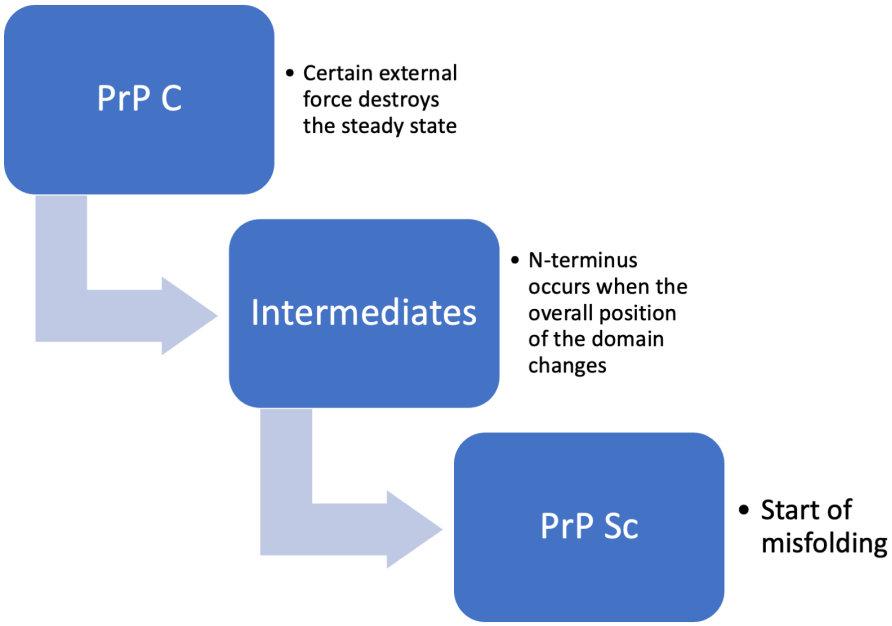

Figure 3.6: The proposed pipeline about how PrP C turns into PrP Sc.

Chapter 4

Module Generation & Methodology

Computer simulations are quite useful tools for neurodegenerative diseases, but their diversity and complexity make their use challenging. Thus, a description and explanation of the conditions and settings specific to the simulation is necessary. This chapter covers the various simulation means, codes, and parameters used in this dissertation. We explain the reasons for the chosen approach even though it may not be the only choice. The setting of the codes and methods are also listed in this chapter.

4.1 Protein Structures and File Formats

Protein is a large biological molecule consisting of one or more long chains of α-amino acid residues linked by peptide bonds. Most proteins naturally fold into a specific three-dimensional structure, and this particular structure is known as the natural state. While most proteins can fold themselves by the nature of their amino acid sequence, many proteins still require the help of molecular chaperones to fold correctly. Under high temperatures or extreme pH conditions, proteins lose their natural structure and activity, a phenomenon

known as denaturation. Biochemists often use the following four aspects to represent the structure of proteins. Protein primary structure: the linear sequence of amino acids that make up the polypeptide chain of a protein. A protein is a polyamide; Protein secondary structure: is a stable structure that relies on hydrogen bonds between C=O and N-H groups between different amino acids, mainly α-helix and β-fold. Because the secondary structure is localized, many regions of other secondary structures can exist in the same protein molecule; Protein tertiary structure: the three-dimensional structure of a protein molecule formed by the arrangement of multiple secondary structure elements in three-dimensional space, which is the overall shape of a single protein molecule. Most tertiary structures of proteins have a hydrophobic core to stabilize the structure, salt bridges (proteins), hydrogen and disulfide bonds, and even post-translational modifications. The term "tertiary structure" can often be expressed as "folding." Tertiary structures control the essential functions of proteins; Protein quaternary structure: A structure formed by several protein molecules (polypeptide chains), often called protein subunits, that function as a protein complex.(Berman et al., 2007; Burley et al., 2017; Sussman et al., 1998; Westbrook & Fitzgerald, 2003)

The key to protein function is the ability to bind specifically and with varying affinity to various other molecules, including protein molecules. The region where proteins bind to other molecules is called the binding site, which is often a "pocket" sunken from the surface of the protein molecule. (Halgren, 2007; Ringe, 1995) The binding ability of a protein is closely related to its tertiary structure, as the structure determines the shape and chemistry of the binding site (i.e., the chemistry of the side chains of amino acid residues surrounding the binding site). Therefore, studying the steric structure of proteins is necessary and meaningful.

4.1.1 Macro Molecules

The Protein Data Bank (PDB) file format(Westbrook & Fitzgerald, 2003) is a textual file describing the three-dimensional structures of molecules held in the Protein Data Bank. The PDB format describes and annotates protein and nucleic acid structures, including atomic coordinates, secondary structure assignments, and atomic connectivity. In addition, experimental metadata are stored. Those files can be seen with a name extension including ".pdb", ".ent", and ".brk". PDB is the legacy file format for the Protein Data Bank, which now keeps data on biological macromolecules in the newer macromolecular Crystallographic Information File (mmCIF) file format.

Both of these file formats are still in use, and their primary purpose is to store essential protein biology information. Therefore, fundamentally, it makes little difference which file format is chosen for this dissertation. Unless otherwise specified, the protein structure files mentioned herein are in PDB format.

4.1.2 Micro Molecules

In most protein files, the protein is not the only content, and the structure often contains several small molecules, which include water molecules and metal ions, in addition to many compound molecules, which are often referred to as ligands for that protein, and the protein is the corresponding receptor for it. These files can be stored as PDB files, but to obtain other molecular structures externally, other file formats will be used. Structure-Data File (SDF) is the one we use for this dissertation. Most of the micromolecules were acquired from $(https://pubchem.ncbi.nlm.nih.gov/)$. (Kim et al., 2019; Kim et al., 2016; Wang et al., 2009)

4.2 Protein-Ligand Docking

Protein-protein docking predicts a complex's structure given a single protein's structure. (Huang & Zou, 2010; Jain, 2006; Sousa et al., 2006) Central to the docking approach is the concept of steric and physicochemical complementarity at the protein-protein interface. In this thesis, all protein-ligand docking was done by Maestro. (Maestro, 2020)

4.2.1 Pre-Process

The structures acquired from online databases need adjustment to maintain the usual processing standard. Directly downloaded PDB files may contain many problems, including unwanted small molecules, missing leading or branched chains of proteins, overlapping protein structures, and file preparation errors. Pre-process is designed to solve these problems. For the protein structures mainly used in this paper, there is no need to repair the main chain of the protein itself, so we only consider the repair of the branched chains. At the same time, we will also remove other solvent molecules at the water molecule level from the structure during the pretreatment process and some small molecular compounds that are irrelevant to the experiment. This is necessary because sometimes these small molecules occupy the original binding site, spatially blocking the optimal compulsory mode and thus leading to suboptimal docking results.

4.2.2 Optimization and Minimization

After getting the non-problematic protein structure, we need to optimize the hydrogen bonding network. The tasks were achieved by Optimize in Protein Preparation Wizard, which includes optimizing the orientation of hydroxyl groups, performing 180° flips of the amide groups of asparagine and glutamine and the ring of a histidine, and adjusting the charge state of histidine residues.

The job then orients water molecules to optimize hydrogen bonding. This iterative process passes over all the groups whose H-bonds must be optimized multiple times. Subsequently, minimization is used as a final step to optimize the entire protein structure, a restrained process that does not change the protein structure to a great extent.

4.2.3 Ligand Preparation

Ligands usually refer to some small molecule compound. In experiments to achieve protein-ligand docking, ligands can often be divided into original ligands and external ligands. For the original ligands, which are included in the PDB file, we separate them out and store them as a separate section, while for the external ligands, we usually obtain the SDF file from the PubChem website and import it into Maestro. We use LigPrep to process individual ligand files to generate scientifically sound molecular models to enumerate the different structures that ligands can sample, as these variations can lead to very different results in subsequent calculations. A generic conversion program configured to generate a library of ligands with the desired structural and chemical characteristics can significantly simplify the entire computer drug discovery process. Here we have chosen to output five candidate structural forms, keeping only the highest-scoring frame.

4.2.4 Receptor Grid Generation

Glide (Grid-based Ligand Docking with Energetics) searches for favorable interactions between one or more typically small ligand molecules and a larger receptor molecule, usually a protein. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. The Receptor Grid Generation panel contains five tabs, which are: Receptor tab - use this tab to define the re-

ceptor (by identifying the Workspace ligand, if present) and optionally to scale the van-der-Waals radius of the receptor atom; Site tab - use this tab to determine the volume for which the Use this tab to select the position and size of the volume for which the lattice will be generated, representing the active sites of the receptor; Constraints tab - use this tab to specify certain receptor atoms for ligand-receptor interactions (position/NOE, H-bond/metal constraints), which you can then choose to require during docking work; Rotatable groups - use this tab to specify the groups that should be considered rotatable in the lattice generation groups that should be regarded as rotatable in the lattice generation. These groups are currently limited to the hydroxyl groups in Ser, Thr and Tyr, and the thiols in Cys. Excluded Volumes - Use this tab to set the spatial regions where the ligands are penalized during docking.

4.2.5 Main Docking

HTVS and SP docking use the same scoring function. However, HTVS reduces the number of intermediate conformations throughout the docking funnel and the thoroughness of final twist refinement and sampling. The docking algorithm itself is the same. XP does more extensive sampling than SP: it starts with SP sampling before its anchoring and growth procedures. XP also employs a more complex scoring function that is "harder" than SP's GlideScore, requiring more complementarity of ligand-receptor shapes. This rules out the false positives allowed by SP. Because XP can penalize inappropriate ligands for using a specific receptor conformation, we recommend docking multiple receptor conformations if possible.

4.2.6 MM-GBSA Analysis

Prime MM-GBSA generates a lot of energy properties. These properties report energies for the ligand, receptor, and complex structures, as well as energy

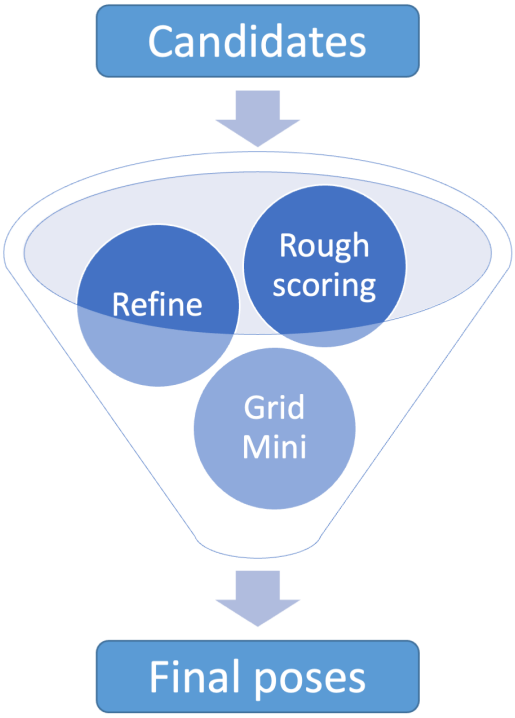

Figure 4.1: Protein-Ligand Docking process Funnel

differences relating to strain and binding, and are broken down into contributions from various terms in the energy expression. There are five fundamental energy calculations done in Prime MM-GBSA: Optimized free receptor (="Receptor") Optimized free ligand (="Ligand") Optimized complex (="Complex") Receptor from minimized/optimized complex Ligand from minimized/optimized complex From these energies, the following strain and binding energies are calculated:

$$Rec\,Strain = Receptor(from\,optimized\,complex) - Receptor \quad (4.1)$$

$$Lig\ Strain = Ligand(from\ optimized\ complex) - Ligand \tag{4.2}$$

$$MMGBSA\ dG\ Bind = Complex - (Receptor + Ligand) \tag{4.3}$$

$$\begin{aligned} MMGBSA\ dG\ Bind(NS) = Complex - (Receptor(optimized\ complex) \\ + Ligand(optimized\ complex)) \\ = MMGBSA\ dG\ Bind - (Rec\ Strain + Lig\ Strain) \end{aligned} \tag{4.4}$$

The MM-GBSA score reflects the strength of the binding capacity, which is usually a negative value, with larger absolute values implying a more robust binding degree.

4.2.7 Protein-protein Docking

The process of protein-protein docking differs significantly from traditional protein-ligand docking in that it requires consideration of both the possible structural changes and the mode of interaction between the receptor protein and the ligand protein. GRAMM-X and PyDock are the 2 tools that we used in this dissertation for conducting protein-protein docking. (Cheng et al., 2007; Pallara et al., 2017; Rosell et al., 2020; Tovchigrechko & Vakser, 2005; Tovchigrechko & Vakser, 2006)

4.3 Molecule Dynamic Simulations

Molecular dynamics (MD) is a computer simulation method for analyzing the physical motion of atoms and molecules. It allows atoms and molecules to interact with each other for a fixed period of time and thus to understand the dynamic "evolution" of the system. In the most common version, the trajectories of atoms and molecules are determined by numerically solving the Newtonian equations of motion for a system of interacting particles, where the forces between the particles and their potential energies are usually calculated using the

interatomic potential or the molecular mechanic's force field. Since molecular systems usually consist of a large number of particles, the properties of such complex systems cannot be determined analytically; MD simulations bypass this problem by using numerical methods

4.3.1 GROMACS

GROMACS is a free, open-source software suite for high-performance molecular dynamics and output analysis. GROMACS operates through a command line interface and can use files for input and output. It provides feedback for calculating progress and estimated time of arrival (ETA), a trajectory viewer, and a comprehensive library for trajectory analysis. In addition, the support for different force fields makes GROMACS very flexible. It can be executed in parallel using a Message Passing Interface (MPI) or threads. It contains a script for converting molecular coordinates in a Protein Data Bank (PDB) file into the format it uses internally. Once a configuration file for the simulation of several molecules (possibly including the solvent) has been created, the simulation run (which can be time-consuming) generates a trajectory file describing the motion of the atoms over time. This file can then be analyzed or visualized using several provided tools.

4.3.2 Amber

Amber is a suite of bio-molecular simulation programs. Amber is distributed in two parts: AmberTools22 and Amber22. AmberTools22 is mainly used to prepare files for Amber22 to run the central MD. For the system with only proteins, we directly use tLEaP to preprocess the proteins, generate the coordinate files and topology files required for the simulation, and store them in two files, prm7 and rst7, respectively; the force fields were used to describe the poten-

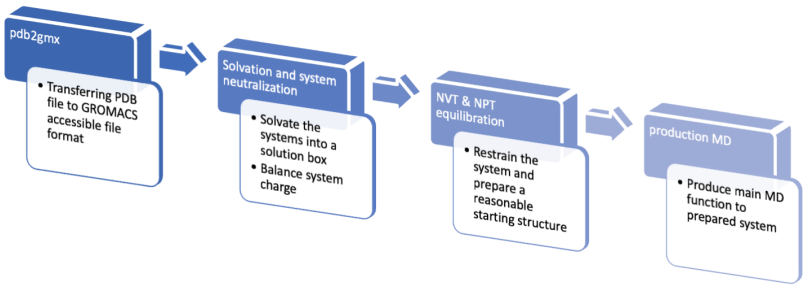

Figure 4.2: The summarized process of GROMACS MD simulation pipeline for a pure protein system. The main function was done in Linux condition with GPU

tial energy 4.1; while for the system with proteins and ligands, Antechamber package is used to process the ligand files, and then the proteins are processed almost indistinguishably from before, and finally, the information of proteins and ligands are combined together and stored in two files, prm7 and rst7, for subsequent operations.

Table 4.1: Force-fields that were used in AmberTools22 to process PDB files

Forcefield	Forcefield applicable subject
leaprc.protein.ff14SB	Protein molecules
leaprc.water.tip3p	Water molecules
leaprc.gaff	Chemical compounds/Ligands

The Amber commands that were used in this dissertation are listed below: Commands that load the package and make the running condition:

```
$ ml Amber/18-fosscuda-2018b-AmberTools-18-patchlevel-
```

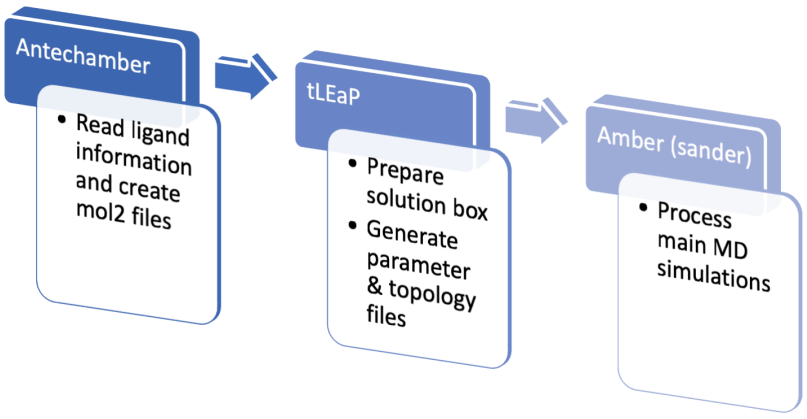

Figure 4.3: The summarized process of Amber MD simulation pipeline. The Antechamber package is optional only if the system consists of ligand components. tLEaP in the AmberTools22 package was done to create the necessary parameter and topology files for the main MD function. The main function was done in Linux condition with GPU

```
$ export PYTHONNOUSERSITE=1
```

Commands that prepare pure proteins:

```
$ pdb4amber -i input.pdb -o output.pdb --reduce --dry
$ tleap
    - source leaprc.protein.ff14SB
    - source leaprc.GLYCAM_06j-1
    - source leaprc.water.tip3p
    - source leaprc.RNA.OL3
    ----------------------------------------
    - unit = loadPDB output.pdb
    - charge unit
    - addIons unit Na+ x [Cl- y]
    - solvatebox unit TIP3PBOX 10.0
    - saveamberparm unit prm7 rst7
```

Commands that prepare proteins with ligands:

```
$ antechamber -i XXX.mol2 -fi mol2 -o XXX.gaff.mol2 -fo mol2
$ parmchk2 -i XXX.gaff.mol2 -f mol2 -o XXX.ante.frcmod
$ pdb4amber -i input.pdb -o output.pdb --reduce --dry
$ tleap
    - source leaprc.protein.ff14SB
    - source leaprc.GLYCAM_06j-1
    - source leaprc.water.tip3p
    - source leaprc.RNA.OL3
    ----------------------------------------
    - REC = loadpdb output.pdb
    - loadamberparams XXX.ante.frcmod
    - XXX = loadmol2 XXX.gaff.mol2
    - COM = combine {REC XXX}
    - charge unit
    - addIons unit Na+ x [Cl- y]
    - solvatebox COM TIP3PBOX 10.0
    - charge COM
    - addIons COM Na+ n
    - saveamberparm unit prm7 rst7
```

After preparing the solution box files, the central MD was calculated using GPU, with the resources provided by GACRC at UGA and XSEDE. We used bash commands to run these jobs and the commands are listed below:

```
$ AMBERHOME/bin/pmemd.cuda -O -i 01_Min.in
-o 01_Min.out -p prm7-c rst7 -r 01_Min.ncrst
-inf 01_Min.mdinfo
$ AMBERHOME/bin/pmemd.cuda -O -i 02_Heat.in
-o 02_Heat.out -p prm7-c 01_Min.ncrst -r 02_Heat.ncrs
-x 02_Heat.nc -inf 02_Heat.mdinfo
$ AMBERHOME/bin/pmemd.cuda -O -i 03_Prod.in
-o 03_Prod.out -p prm7 -c 02_Heat.ncrst
-r 03_Prod.ncrst -x 03_Prod.nc -inf 03_Prod.info
```

4.3.3 CHARMM

CHARMM is a molecular simulation program with broad application to many-particle systems with a comprehensive set of energy functions, enhanced sampling methods, and support for multi-scale techniques, including QM/MM, MM/CG, and a range of implicit solvent models. CHARMM primarily uses to generate necessary topology and parameter files for ligands and glycans and then transfers them to GRAMACS readable files.

CHARMM-GUI was used for transferring jobs.(Jo et al., 2008a) When we work with protein-ligand conjugates, we first isolate the structure of the ligand into a separate file (.pdb or .sdf), and then we use Input Generator to read this ligand file and use Glycan Reader& Modeler or Ligand Reader & Modeler to process it and get the necessary topology file and coordinate file. (Jo et al., 2008b; J. Lee et al., 2016)

4.3.4 Post simulation analysis

After the MD, we used VMD to read the simulation-generated files to generate frame-by-frame dynamic model changes to analyze their trajectory changes and

spatial position migration. Since the determination of secondary structure is essential for this thesis and stereo structural analysis is essential, we used the sscache package to obtain real-time dynamic secondary structure changes. The Following codes were loaded to tk console, stored with the name of sscache, and called by the command "start_sscache"

```
proc start_sscache {{molid top}} {
    global sscache_data
    if {! [string compare $molid top]} {
        set molid [molinfo top]
    }
    global vmd_frame
    return
}
proc stop_sscache {{molid top}} {
    if {! [string compare $molid top]} {
        set molid [molinfo top]
    }
    global vmd_frame
    trace vdelete vmd_frame($molid) w sscache
    return
}
```

```
proc reset_sscache {} {
    global sscache_data
    if [info exists sscache_data] {
        unset sscache_data
    }
    return
}
proc sscache {name index op} {
    global sscache_data
    set sel [atomselect $index "protein name CA"]
    set frame [molinfo $index get frame]
    if [info exists sscache_data($index,$frame)] {
        $sel set structure $sscache_data($index,$fram
        return
    }
    vmd_calculate_structure $index
    set sscache_data($index,$frame) [$sel get structu
    return
}
```

4.4 Protein structure prediction

The biological functions of proteins are highly correlated with their stereo structures. The same protein sequence under different steric folds can have different biological properties. Existing protein databases (e.g., RCSB) store data on numerous protein structures, but many unlisted proteins exist. The common means of obtaining protein structures is to irradiate the protein in its crystalline state with the help of X-rays or to probe the main domains with the help of

NMR, so there are many protein individuals lose their N-terminus solid structures through that way

4.4.1 RoseTTAFold & Robetta

All the protein structure prediction models in this dissertation were single-chain modeling. The primary use of the RoseTTAFold algorithm in this experiment is through the Robetta server. ($https://robetta.bakerlab.org/$) RoseTTAFold was used for more accurate modeling methods by applying deep learning. The comparative modeling and Ab inito options are not selected because no sufficient models are available in the Protein Data Bank by the time we edit this dissertation.

4.4.2 AlphaFold

We used AlphaFold with the help of the Georgia Advanced Computing Resource Center(GACRC). AlphaFold is open-source software; its source code GitHub address is: ($https://github.com/deepmind/alphafold$) We set up the environment according to the instructions in the readme file on GitHub. (Evans et al., 2021; Jumper et al., 2021) AlphaFold 2.0.0 was used for all structure predictions in this dissertation. The bash file was used for the main prediction. Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Chapter 5

Conclusion and Future Plan

Neurodegenerative diseases pose a significant threat to humanity now and in the future, and research on neurodegenerative diseases is urgently needed. Due to the unknown and specific nature of the pathology of these diseases, it has become essential to understand the particular factors that can trigger structural changes in distinct proteins. In this dissertation, computer technology is introduced as a research tool to explore the pathology of neurodegenerative diseases. Through molecular dynamics simulations for specific proteins, we have successfully reproduced the change in the initiation state of protein misfolding.

In Chapter 2, we examine Alzheimer's disease. For the monomeric Aβ40, we analyzed its structural properties. We constructed a dimer using the Aβ that has been misfolded with the normal Aβ. We have performed several computer simulation studies using this model. The experimental results reveal the misfolding initiation process of normal Aβ: a new beta-sheet is generated in a specific amino acid region and is stably retained without vanishing. Subsequently, a surface charge assay of the dimer revealed a significant difference in charge of the dimer at amino acids 15-17 under acidic conditions, while the predicted binding site of the dimer showed that the protein has the highest potential to bind small

molecules at the same position. Based on this discovery, and in conjunction with previous literature, we attempted to bind heparin at this active site. The results of molecular dynamics simulation experiments showed that heparin has a significant ability to maintain a dimeric parallel structure, which provides the possibility that Aβ can aggregate layer by layer. Subsequently, we considered how to generate the first misfolded structure and tried other types of proteins to guide the possibility of misfolding of Aβ. We found that glycoprotein B of the Herpes Simplex Virus type 1(HSV-1) could achieve this effect through experiments. HSV-1 Glycoprotein B can trigger secondary structure changes in Aβ at the exact location and is even more stable than Aβ dimer in a similar time frame as the simulation experiments with Aβ dimer.

In chapter 3, We conducted a similar study on prions. We found the possibility of similar secondary structure changes in critical proteins of prions. PrP has a more extended sequence than Aβ, while its steric structure is incomplete in the RCSB database, and its N-terminal structure is unknown. The available evidence supports the possibility of a region at the N-terminal end of PrP that determines the global protein secondary structure, and therefore structural prediction for the N-terminal end is critical for experiments. We used AlphaFold to process the protein sequence of PrP to obtain a full-length model of PrP. Simulations for the PrP structure reveal that amino acid positions 162-163 are essential for determining the onset of changes in its secondary structure. At the same time, we find that the position of the N end of the PrP relative to the primary domain is critical. Based on the results of existing simulations and previous studies, we propose a novel hypothesis for the process of PrP C transformation to PrP Sc. This hypothesis emphasizes the importance of the N-terminus and the possible spatial position transitions that may exist after the steady state of its spatial structure is disrupted by external forces, eventually leading to the beginning of misfolding.

In chapter 4, The various simulation tools, codes, and parameters used in this

dissertation are listed and explained in detail. According to different situations and conditions, we selected the most suitable software tools for the simulation experiments mentioned in Chapters 2 and 3. Specific packages were modified and applied to the simulation to fit our model. Also, this section shows the process of using various simulation programs and lists the command lines used in detail.

With the condition of our current progress, our future work will be aimed at addressing these problems:

- In AD studies, the misfolded infection process of Aβ is currently only at the one-to-one level, and the infection continuum needs to be revealed. the principle of action of Heparin is unclear, and its effect of stabilizing Aβ needs further validated and expanded. simulations of the complete model of HSV-1 Glycoprotein B are desired, which poses a massive challenge for arithmetic power.

- In Human Prion Disease, further studies on the structure of the N-terminus are needed, which include the search for the structure of the stable state, the search for external factors that break the stable state, and the search for the reasons for its spatial location migration.

To sum up, in this dissertation, computational simulation was applied to reveal the pathology of AD and Human Prion Disease. This work has great potential to bring ND research to the next level and provides inspiration to further research for ND treatments.

Appendix A

2

A.1 Distance between Specific atoms in Amino Acids from Aβ and HSV-1

Distance between atoms that may form H-bond are measured. The purpose of this section is to identify whether the interaction between HSV-1 glycoprotein B and Aβ. The captions of the sub figures represents the Atom in the number of amino acid. For instance, in A.1, the distance is between the Oxygen in the 36th amino acids of Aβ and the Nitrogen in the 676th amino acids of HSV-1 glycoprotein B.

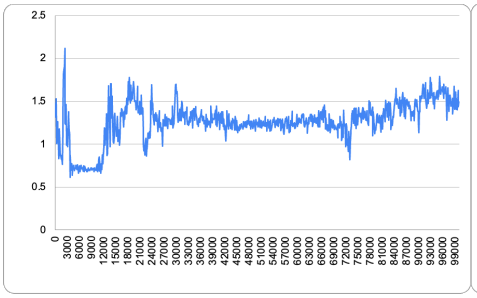

Figure A.1: 36O-676N

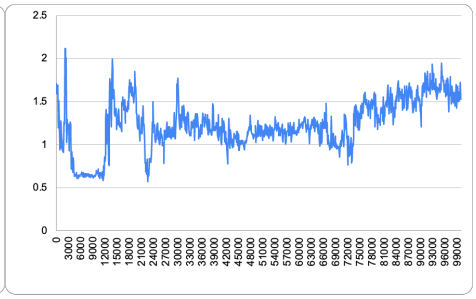

Figure A.2: 36O-677N

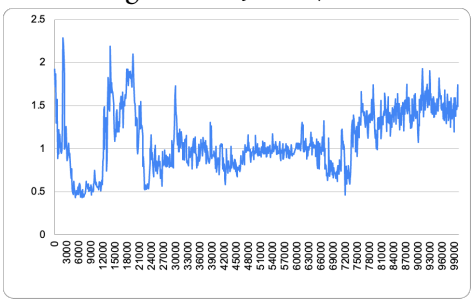

Figure A.3: 36O-678N

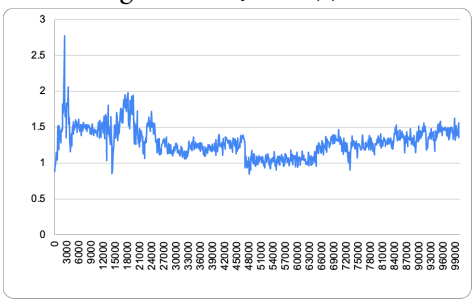

Figure A.4: 36O-673N

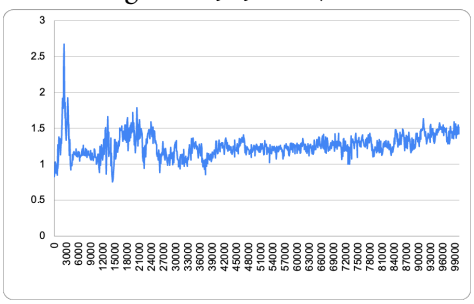

Figure A.5: 37N-673O

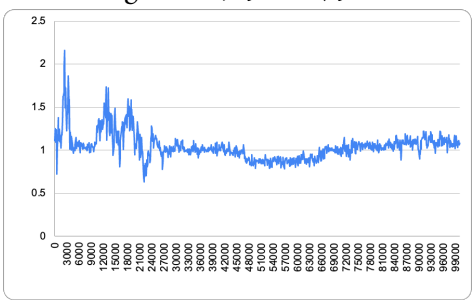

Figure A.6: 36N-674O

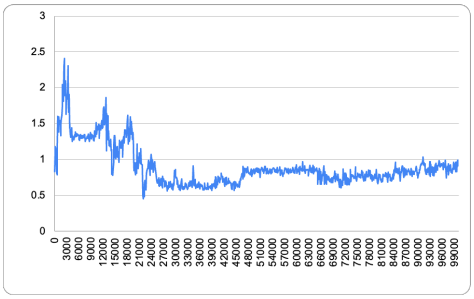

Figure A.7: 35O-674N

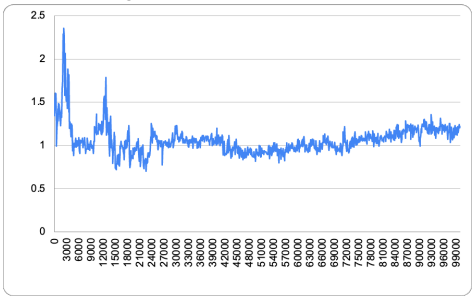

Figure A.8: 35N-675O

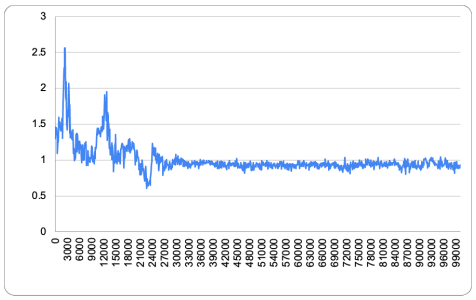

Figure A.8: 34O-675N

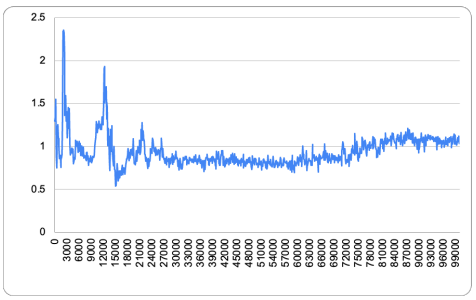

Figure A.9: 34N-676O

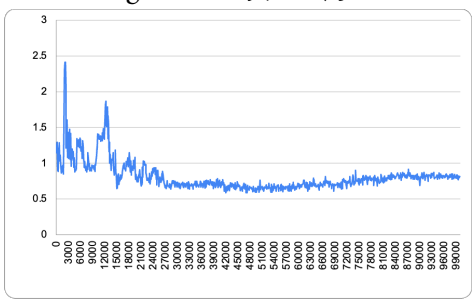

Figure A.10: 33O-676N

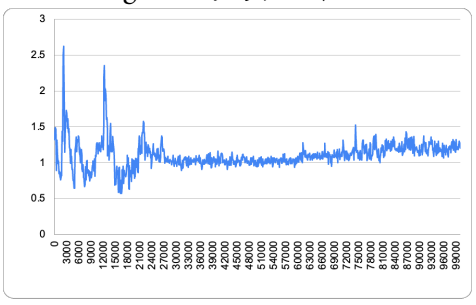

Figure A.11: 33N-677O

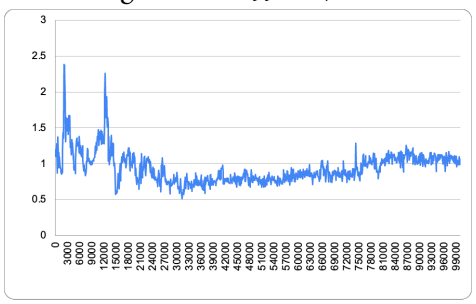

Figure A.12: 32O-677N

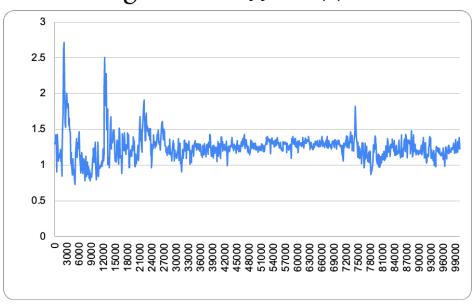

Figure A.13: 32N-678O

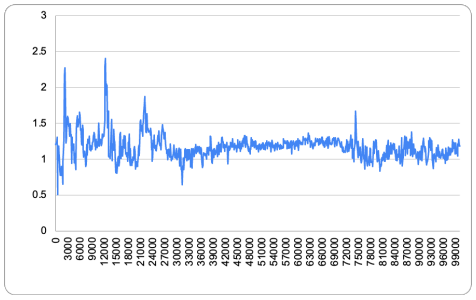

Figure A.14: 31O-678N

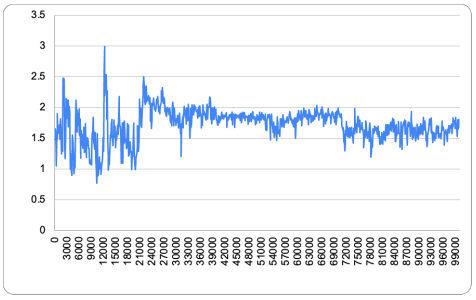

Figure A.15: 31N-679O

Appendix B

3

B.1 Prion protein models MD simulations frame change

The typical frames from the MD simulation of RoseTTA-generated PrP models are listed in Fig.B.1 and Fig.B.2. In both of the figures, graphs A, B, C, and D are sorted according to time. By comparing these frames, we can clearly find that model 1 tends to have consist length of beta sheets, and the whole structure is not producing any obvious secondary change; while for model 2, MD results produced several newly generated beta sheets, which are regarded as the signal of potential misfolding starts.

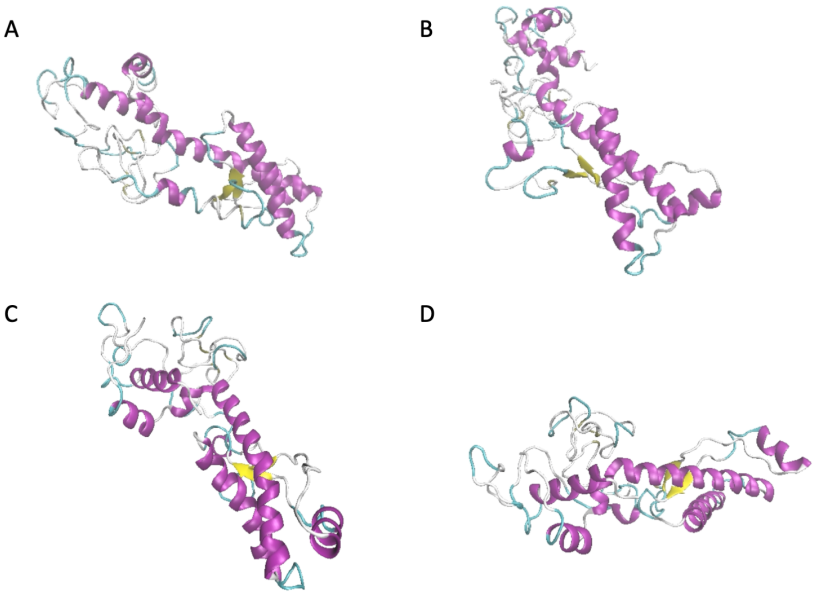

Figure B.1: The typical frames from MD simulation of RoseTTA generated PrP model 1. The

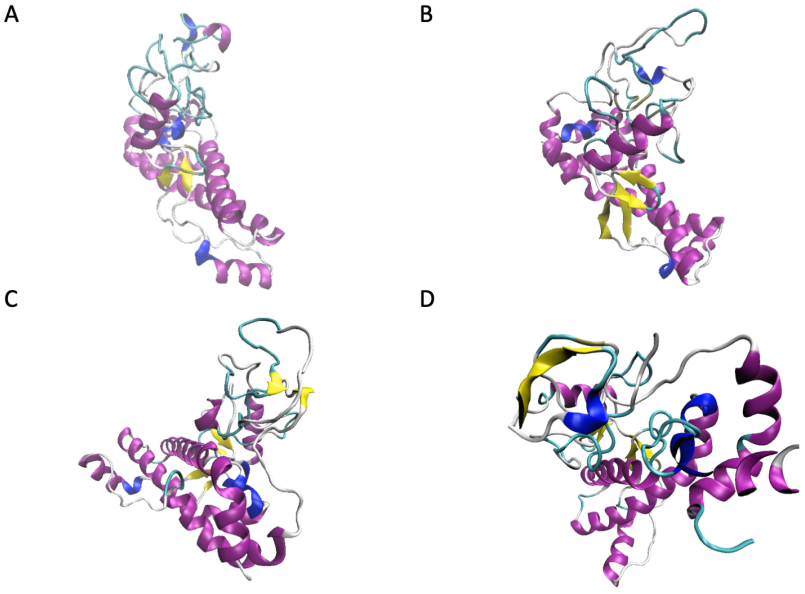

Figure B.2: The typical frames from MD simulation of RoseTTA generated PrP model 2. The

Bibliography

(n.d.). https://www.cdc.gov/prions/index.html

(n.d.). https://www.brightfocus.org/alzheimers/article/history-alzheimers-disease

Abbott, A. (2010). Healthy prions protect nerves. *Nature*. https://doi.org/10.1038/NEWS.2010.29

Alonso, H., Bliznyuk, A. A., & Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, *26*, 531–568. https://doi.org/10.1002/MED.20067

Barz, B., Liao, Q., & Strodel, B. (2018). Pathways of amyloid-β aggregation depend on oligomer shape. *Journal of the American Chemical Society*, *140*(1), 319–327.

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice*, *98*(6), 236–238.

Berman, H., Henrick, K., Nakamura, H., & Markley, J. L. (2007). The worldwide protein data bank (wwpdb): Ensuring a single, uniform archive of pdb data. *Nucleic acids research*, *35*(suppl_1), D301–D303.

Biljan, I., Giachin, G., Ilc, G., Zhukov, I., Plavec, J., & Legname, G. (2012). Structural basis for the protective effect of the human prion protein carrying the dominant-negative e219k polymorphism. *Biochemical Journal*, *446*(2), 243–251.

Blundell, T. L., Sibanda, B. L., Montalvão, R. W., Brewerton, S., Chelliah, V., Worth, C. L., Harmer, N. J., Davies, O., & Burke, D. (2006). Structural

biology and bioinformatics in drug design: Opportunities and challenges for target identification and lead discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 413–423. https://doi.org/10.1098/RSTB.2005.1800

Broe, G., Henderson, A., Creasey, H., McCusker, E., Korten, A., Jorm, A., Longley, W., & Anthony, J. (1990). A case-control study of alzheimer's disease in australia. *Neurology*, *40*(11), 1698–1698.

Brown, P., et al. (1998). 1755 and all that: A historical primer of transmissible spongiform encephalopathy. *Bmj*, *317*(7174), 1688–1692.

Brown, R. C., Lockwood, A. H., & Sonawane, B. R. (2005). Neurodegenerative diseases: An overview of environmental risk factors. *Environmental Health Perspectives*, *113*, 1250–1256. https://doi.org/10.1289/EHP.7567

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., & Velankar, S. (2017). Protein data bank (pdb): The single global macromolecular structure archive. *Protein Crystallography*, 627–641.

Burns, A., & Iliffe, S. (2009). Alzheimer's disease. *BMJ*, *338*(feb05 1), b158–b158. https://doi.org/10.1136/bmj.b158

Checkoway, H., Lundin, J. I., & Kelada, S. N. (2011). Neurodegenerative diseases. *IARC Scientific Publications*, 407–419. https://doi.org/10.1071/rdv24n1iab251

Cheng, T. M.-K., Blundell, T. L., & Fernandez-Recio, J. (2007). Pydock: Electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, *68*(2), 503–515.

Chesebro, B. (2003). Introduction to the transmissible spongiform encephalopathies or prion diseases. *British medical bulletin*, *66*(1), 1–20.

Consortium, U. (2015). Uniprot: A hub for protein information. *Nucleic acids research*, *43*(D1), D204–D212.

Das, R., & Baker, D. (2008). Macromolecular modeling with rosetta. *Annu. Rev. Biochem.*, *77*, 363–382.

Daskalov, A., & Saupe, S. J. (2015). As a toxin dies a prion comes to life: A tentative natural history of the [het-s] prion. *Prion*, *9*(3), 184–189.

De Gage, S. B., Moride, Y., Ducruet, T., Kurth, T., Verdoux, H., Tournier, M., Pariente, A., & Bégaud, B. (2014). Benzodiazepine use and risk of alzheimer's disease: Case-control study. *Bmj*, *349*, g5205.

Dextera, D. T., & Jenner, P. (2013). Parkinson disease: From pathology to molecular disease mechanisms. *Free Radical Biology and Medicine*, *62*, 132–144. https://doi.org/10.1016/J.FREERADBIOMED.2013.01.018

Doecke, J. D., Pérez-Grijalba, V., Fandos, N., Fowler, C., Villemagne, V. L., Masters, C. L., Pesini, P., & and, M. S. (2020). Total asub42/sub/asub40/subratio in plasma predicts amyloid-PET status, independent of clinical AD diagnosis. *Neurology*, *94*(15), e1580–e1591. https://doi.org/10.1212/wnl.0000000000009240

Drug repurposing: Advantages and key approaches | technology networks. (n.d.). https://www.technologynetworks.com/drug-discovery/articles/drug-repurposing-advantages-and-key-approaches-344261

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2021). Protein complex prediction with alphafold-multimer. *bioRxiv*. https://doi.org/10.1101/2021.10.04.463034

Féraudet, C., Morel, N., Simon, S., Volland, H., Frobert, Y., Créminon, C., Vilette, D., Lehmann, S., & Grassi, J. (2005). Screening of 145 anti-prp monoclonal antibodies for their capacity to inhibit prpsc replication in infected cells. *Journal of Biological Chemistry*, *280*(12), 11247–11258.

Finder, V. H., & Glockshuber, R. (2007). Amyloid-β aggregation. *Neurodegenerative Diseases*, *4*(1), 13–27.

Forman, M. S., Trojanowski, J. Q., & Lee, V. M. (2004). Neurodegenerative diseases: A decade of discoveries paves the way for therapeutic breakthroughs. *Nature Medicine 2004 10:10*, *10*, 1055–1063. https://doi.org/10.1038/nm1113

Ghiselli, A., Nardini, M., Baldi, A., & Scaccini, C. (1998). Antioxidant activity of different phenolic fractions separated from an italian red wine. *Journal of agricultural and food chemistry*, *46*(2), 361–367.

Hajduk, P. J., & Greer, J. (2007). A decade of fragment-based drug design: Strategic advances and lessons learned. *Nature Reviews Drug Discovery 2007 6:3*, *6*, 211–219. https://doi.org/10.1038/nrd2220

Halgren, T. (2007). New method for fast and accurate binding-site identification and analysis. *Chemical biology & drug design*, *69*(2), 146–148.

Hane, F., & Leonenko, Z. (2014). Effect of metals on kinetic pathways of amyloid-β aggregation. *Biomolecules*, *4*(1), 101–116.

Huang, S.-Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking. *International journal of molecular sciences*, *11*(8), 3016–3034.

Hung, C. W., Chen, Y. C., Hsieh, W. L., Chiou, S. H., & Kao, C. L. (2010). Ageing and neurodegenerative diseases. *Ageing Research Reviews*, *9*, S36–S46. https://doi.org/10.1016/J.ARR.2010.08.006

Hunter, S., Martin, S., & Brayne, C. (2016). The APP proteolytic system and its interactions with dynamic networks in alzheimer's disease. In *Systems biology of alzheimer's disease* (pp. 71–99). Springer New York. https://doi.org/10.1007/978-1-4939-2627-5_3

Hurd, M. D., Martorell, P., Delavande, A., Mullen, K. J., & Langa, K. M. (2013). Monetary costs of dementia in the united states [PMID: 23550670]. *New England Journal of Medicine*, *368*(14), 1326–1334. https://doi.org/10.1056/NEJMsa1204629

Ireland, N. (2003). Bovine spongiform encephalopathy.

Jain, A. N. (2006). Scoring functions for protein-ligand docking. *Current Protein and Peptide Science*, *7*(5), 407–420.

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008a). Charmm-gui: A web-based graphical user interface for charmm. *Journal of computational chemistry*, *29*(11), 1859–1865.

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008b). Charmm-gui: A web-based graphical user interface for charmm. *Journal of computational chemistry*, *29*(11), 1859–1865.

Juergenliemk, G., Boje, K., Huewel, S., Lohmann, C., Galla, H.-J., & Nahrstedt, A. (2003). In vitro studies indicate that miquelianin (quercetin 3-o-ß-d-glucuronopyranoside) is able to reach the cns from the small intestine. *Planta medica*, *69*(11), 1013–1017.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaufmann, K. W., Lemmon, G. H., DeLuca, S. L., Sheehan, J. H., & Meiler, J. (2010). Practically useful: What the rosetta protein modeling suite can do for you. *Biochemistry*, *49*(14), 2987–2998.

Khalili-Shirazi, A., Summers, L., Linehan, J., Mallinson, G., Anstee, D., Hawke, S., Jackson, G. S., & Collinge, J. (2005). Prp glycoforms are associated in a strain-specific ratio in native prpsc. *Journal of General Virology*, *86*(9), 2635–2644.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., et al. (2019). Pubchem 2019 update:

Improved access to chemical data. *Nucleic acids research*, *47*(D1), D1102–D1109.

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., et al. (2016). Pubchem substance and compound databases. *Nucleic acids research*, *44*(D1), D1202–D1213.

Kovacs, G. G., & Budka, H. (2008). Prion diseases: From protein to cell pathology. *The American Journal of Pathology*, *172*, 555–565. https://doi.org/10.2353/AJPATH.2008.070442

Lee, G., Park, C., & Ahn, J. (2019). Novel deep learning model for more accurate prediction of drug-drug interaction effects. *BMC Bioinformatics*, *20*. https://doi.org/10.1186/S12859-019-3013-0

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., et al. (2016). Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Journal of chemical theory and computation*, *12*(1), 405–413.

Linden, R., Martins, V. R., Prado, M. A., Cammarota, M., Izquierdo, I., & Brentani, R. R. (2008). Physiology of the prion protein. *Physiological reviews*, *88*(2), 673–728.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, *2012*(251364), 251364.

Luo, Y., Smith, J. V., Paramasivam, V., Burdick, A., Curry, K. J., Buford, J. P., Khan, I., Netzer, W. J., Xu, H., & Butko, P. (2002). Inhibition of amyloid-β aggregation and caspase-3 activation by the ginkgo biloba extract egb761. *Proceedings of the national academy of sciences*, *99*(19), 12197–12202.

Ma, J., & Lindquist, S. (2002). Conversion of prp to a self-perpetuating prpsc-like conformation in the cytosol. *Science*, *298*(5599), 1785–1788.

Maestro, S. (2020). Maestro. *Schrödinger, LLC, New York, NY*, *2020*.

Mandaci, S. Y., Caliskan, M., Sariaslan, M. F., Uversky, V. N., & Coskuner-Weber, O. (2020). Epitope region identification challenges of intrinsically disordered proteins in neurodegenerative diseases: Secondary structure dependence of α-synuclein on simulation techniques and force field parameters. *Chemical Biology & Drug Design*, *96*, 659–667. https://doi.org/10.1111/CBDD.13662

Mandal, S., Moudgil, M., & Mandal, S. K. (2009). Rational drug design. *European Journal of Pharmacology*, *625*, 90–100. https://doi.org/10.1016/J.EJPHAR.2009.06.065

Matthews, K. A., Xu, W., Gaglioti, A. H., Holt, J. B., Croft, J. B., Mack, D., & McGuire, L. C. (2018). Racial and ethnic estimates of alzheimer's disease and related dementias in the united states (2015–2060) in adults aged $\geq$65 years. *Alzheimer's &amp Dementia*, *15*(1), 17–24. https://doi.org/10.1016/j.jalz.2018.06.3063

Mohelska, H., Maresova, P., Valis, M., & Kuca, K. (2015). Alzheimer's disease and its treatment costs: Case study in the czech republic. *Neuropsychiatric Disease and Treatment*, *11*, 2349.

Murphy, R. M. (2002). Peptide aggregation in neurodegenerative disease. *Annu. Rev. Biomed. Eng*, *4*, 155–74. https://doi.org/10.1146/annurev.bioeng.4.092801.094202

Ndagi, U., Falaki, A. A., Abdullahi, M., Lawal, M. M., & Soliman, M. E. (2020). Antibiotic resistance: Bioinformatics-based understanding as a functional strategy for drug design. *RSC Advances*, *10*, 18451–18468. https://doi.org/10.1039/D0RA01484B

Nedelec, T., Couvy-Duchesne, B., Monnet, F., Daly, T., Ansart, M., Gantzer, L., Lekens, B., Epelbaum, S., Dufouil, C., & Durrleman, S. (2022). Identifying health conditions associated with alzheimer's disease up to

15 years before diagnosis: An agnostic study of french and british health records. *The Lancet Digital Health*, *4*(3), e169–e178.

Oprea, T. I., Bauman, J. E., Bologa, C. G., Buranda, T., Chigaev, A., Edwards, B. S., Jarvik, J. W., Gresham, H. D., Haynes, M. K., Hjelle, B., Hromas, R., Hudson, L., MacKenzie, D. A., Muller, C. Y., Reed, J. C., Simons, P. C., Smagley, Y., Strouse, J., Surviladze, Z., … Sklar, L. A. (2011). Drug repurposing from an academic perspective. *Drug Discovery Today: Therapeutic Strategies*, *8*, 61–69. https://doi.org/10.1016/J.DDSTR.2011.10.002

Pallara, C., Jiménez-García, B., Romero, M., Moal, I. H., & Fernández-Recio, J. (2017). Pydock scoring for the new modeling challenges in docking: Protein–peptide, homo-multimers, and domain–domain interactions. *Proteins: Structure, Function, and Bioinformatics*, *85*(3), 487–496.

Panigrahi, S., Stetefeld, J., Jangamreddy, J. R., Mandal, S., Mandal, S. K., & Los, M. (2012). Modeling of molecular interaction between apoptin, bcr-abl and crkl - an alternative approach to conventional rational drug design. *PLOS ONE*, *7*, e28395. https://doi.org/10.1371/JOURNAL.PONE.0028395

Park, H., Kim, D. E., Ovchinnikov, S., Baker, D., & DiMaio, F. (2018). Automatic structure prediction of oligomeric assemblies using robetta in casp12. *Proteins: Structure, Function, and Bioinformatics*, *86*, 283–291.

Polo, J. (2000). The history and classification of human prion diseases. *Revista de Neurologia*, *31*(2), 137–141.

Poser, C. M. (2002). Notes on the history of the prion diseases. part i. *Clinical neurology and neurosurgery*, *104*(1), 1–9.

Prieto-Martínez, F. D., López-López, E., Juárez-Mercado, K. E., & Medina-Franco, J. L. (2019). Computational drug design methods—current and future perspectives. *In Silico Drug Design*, 19–44. https://doi.org/10.1016/B978-0-12-816125-8.00002-X

Prusiner, S. B. (1991). Molecular biology of prion diseases. *Science*, *252*(5012), 1515–1522.

Prusiner, S. B., & Hsiao, K. K. (1994). Human prion diseases. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *35*(4), 385–395.

Prusiner, S. B., Scott, M. R., DeArmond, S. J., & Cohen, F. E. (1998). Prion protein biology. *cell*, *93*(3), 337–348.

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2018). Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery 2018 18:1*, *18*, 41–58. https://doi.org/10.1038/nrd.2018.168

Reis-Filho, J. S. (2009). Next-generation sequencing. *Breast cancer research*, *11*(3), 1–7.

Ringe, D. (1995). What makes a binding site a binding site? *Current opinion in structural biology*, *5*(6), 825–829.

Rosell, M., Rodrıguez-Lumbreras, L. A., & Fernández-Recio, J. (2020). Modeling of protein complexes and molecular assemblies with pydock. In *Protein structure prediction* (pp. 175–198). Springer.

Rudrapal, M., Khairnar, S. J., & Jadhav, A. G. (2020). Drug repurposing (dr): An emerging approach in drug discovery. *Drug Repurposing - Hypothesis, Molecular Aspects and Therapeutic Applications*. https://doi.org/10.5772/INTECHOPEN.93193

Scheltens, P., Blennow, K., Breteler, M. M., de Strooper, B., Frisoni, G. B., Salloway, S., & der Flier, W. M. V. (2016). Alzheimer's disease. *Lancet (London, England)*, *388*, 505–517. https://doi.org/10.1016/S0140-6736(15)01124-1

Schilling, K. M., Tao, L., Wu, B., Kiblen, J. T., Ubilla-Rodriguez, N. C., Pushie, M. J., Britt, R. D., Roseman, G. P., Harris, D. A., & Millhauser, G. L.

(2020). Both n-terminal and c-terminal histidine residues of the prion protein are essential for copper coordination and neuroprotective self-regulation. *Journal of molecular biology*, *432*(16), 4408–4425.

Selkoe, D. J. (1991). The molecular pathology of alzheimer's disease. *Neuron*, *6*, 487–498. https://doi.org/10.1016/0896-6273(91)90052-2

Silveira, J. R., Raymond, G. J., Hughson, A. G., Race, R. E., Sim, V. L., Hayes, S. F., & Caughey, B. (2005). The most infectious prion protein particles. *Nature*, *437*(7056), 257–261.

Sosa-Ortiz, A. L., Acosta-Castillo, I., & Prince, M. J. (2012). Epidemiology of dementias and alzheimer's disease. *Archives of medical research*, *43*(8), 600–608.

Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein–ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 15–26.

Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., & Abola, E. E. (1998). Protein data bank (pdb): Database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica Section D: Biological Crystallography*, *54*(6), 1078–1084.

Terreros-Roncal, J., Moreno-Jiménez, E. P., Flor-García, M., Rodríguez-Moreno, C. B., Trinchero, M. F., Cafini, F., Rábano, A., & Llorens-Martín, M. (2021). Impact of neurodegenerative diseases on human adult hippocampal neurogenesis. *Science*, *374*, 1106–1113. https://doi.org/10.1126/SCIENCE.ABL5163/SUPPL_FILE/SCIENCE.ABL5163_DATA_S1_AND_S2.ZIP

Tovchigrechko, A., & Vakser, I. A. (2005). Development and testing of an automated approach to protein docking. *Proteins: Structure, Function, and Bioinformatics*, *60*(2), 296–301.

Tovchigrechko, A., & Vakser, I. A. (2006). Gramm-x public web server for protein–protein docking. *Nucleic acids research*, *34*(suppl_2), W310–W314.

Tuo, W., Zhuang, D., Knowles, D. P., Cheevers, W. P., Sy, M.-S., & O'Rourke, K. I. (2001). Prp-c and prp-sc at the fetal-maternal interface. *Journal of Biological Chemistry*, *276*(21), 18229–18234.

Tyler, K. L., et al. (2004). Herpes simplex virus infections of the central nervous system: Encephalitis and meningitis, including mollaret's. *HERPES-CAMBRIDGE-*, *11*, 57A–64A.

Wadsworth, J. D., & Collinge, J. (2010). Molecular pathology of human prion disease. *Acta Neuropathologica 2010 121:1*, *121*, 69–77. https://doi.org/10.1007/S00401-010-0735-5

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). Pubchem: A public information system for analyzing bioactivities of small molecules. *Nucleic acids research*, *37*(suppl_2), W623–W633.

Westbrook, J. D., & Fitzgerald, P. (2003). The pdb format, mmcif, and other data formats. *Methods Biochem Anal*, *44*, 161–179.

Westergard, L., Christensen, H. M., & Harris, D. A. (2007). The cellular prion protein (prpc): Its physiological function and role in disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1772*(6), 629–644.

Wilesmith, J. W., Wells, G., Cranwell, M. P., & Ryan, J. (1988). Bovine spongiform encephalopathy: Epidemiological studies. *The Veterinary Record*, *123*(25), 638–644.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). Drugbank 5.0: A major update to the

drugbank database for 2018. *Nucleic Acids Research*, *46*, D1074–D1082. https://doi.org/10.1093/NAR/GKX1037

Wollacott, A. M., Robinson, L. N., Ramakrishnan, B., Tissire, H., Viswanathan, K., Shriver, Z., & Babcock, G. J. (2019). Structural prediction of antibody-april complexes by computational docking constrained by antigen saturation mutagenesis library data. *Journal of Molecular Recognition*, *32*, e2778. https://doi.org/10.1002/JMR.2778

Wulf, M.-A., Senatore, A., & Aguzzi, A. (2017). The biological function of the cellular prion protein: An update. *BMC biology*, *15*(1), 1–13.

Ye, J., McGinnis, S., & Madden, T. L. (2006). Blast: Improvements for better sequence analysis. *Nucleic acids research*, *34*(suppl_2), W6–W9.

Yoshiike, Y., Tanemura, K., Murayama, O., Akagi, T., Murayama, M., Sato, S., Sun, X., Tanaka, N., & Takashima, A. (2001). New insights on how metals disrupt amyloid β-aggregation and their effects on amyloid-β cytotoxicity. *Journal of Biological Chemistry*, *276*(34), 32293–32299.

Young, D. C. (2009). Computational drug design : A guide for computational and medicinal chemists, 307.